

Universidade Federal de Santa Catarina
Programa de Pós-Graduação em Engenharia de Produção

**FAN 2002: UM MODELO NEURO-FUZZY PARA
RECONHECIMENTO DE PADRÕES**

Roberto Tadeu Raittz

Tese apresentada ao Programa de Pós-graduação em Engenharia de Produção da Universidade Federal de Santa Catarina como requisito parcial para obtenção do título de Doutor em Engenharia de Produção.

Orientador: Prof. Dr. Fernando Álvaro Ostuni
Gauthier

Florianópolis
2002

Roberto Tadeu Raittz

**FAN 2002: UM MODELO NEURO-FUZZY PARA
RECONHECIMENTO DE PADRÕES**

Esta tese foi julgada e aprovada para a obtenção do título de Doutor em Engenharia de Produção no Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina.

Florianópolis, 12 de agosto de 2002

Prof. Ricardo Miranda Barcia, PhD
Coordenador do Curso

BANCA EXAMINADORA

Prof. Fernando Álvaro Ostuni Gaultier, Dr. - Orientador

Profa. Ana Maria Bencxivani Franzoni Dra.

Prof. Paulo Afonso Bracarense da Costa, Dr.

Prof. Marco A. B. Candido, Dr.

Prof. Malcon Anderson Tafner, Dr.

Prof. Rogério Cid Bastos, Dr.

Prof. Alejandro Martinez, Dr.

Prof. Roberto Carlos Santos Pacheco, Dr.

Dedicado a Rosângela.

AGRADECIMENTOS

Aos meus pais, Alvino Raittz e Jamir Maria Raittz, pela vida e pelos ensinamentos que me permitiram percorrer o caminho da curiosidade e do interesse pelo conhecimento científico.

A minhas irmãs, cunhados – que são como irmãos – e sobrinhos, com amor.

Ao meu orientador, professor doutor Fernando Álvaro Ostoni Gauthier, pelo incentivo.

Ao amigo e professor Paulo Afonso Bracarense da Costa, pela contribuição na consistência do método estabelecido neste trabalho.

Ao professor Roberto Carlos Santos Pacheco, pela força enorme que me foi dada no início do processo de estudos do tema.

Aos professores Alejandro Martines, Édis Mafra Lapoli e José Leomar Todesco, pela colaboração.

Aos matemáticos João Arthur de Souza e Gertrudes Dandolini, pela confiança no processo de desenvolvimento do modelo FAN; pelo acolhimento nas questões acadêmicas e também pelo convívio amigo e solidário.

Aos colegas Pedro Von Mecheln, Mauro José Belli e Luiz Alberto Brettas, pelo estímulo oferecido ao longo desse percurso.

Aos amigos Luiz Ermindo Cavallet, Izabel e Ana Paula Raittz Cavallet, Dieval Guizeline e Jussara PRPPG, pelo apoio.

Às instituições Capes e ET-UFPR, pelo auxílio.

Ao professor, doutor e físico Hamilton de Araújo Bicalho, *in memoriam*, pelos conhecimentos transmitidos.

À jornalista Marlise de Cássia Bassfeld, pelas interlocuções, revisão e edição do texto.

A todos que direta ou indiretamente contribuíram para a concretização deste trabalho, sou, sinceramente,

Muito grato!

יוצד'ק 'סוד עולם
“*O justo é a fundação do mundo.*”

O Bahir 61, provérbios 10,25

SUMÁRIO

	LISTA DE TABELAS	ix
	LISTA DE FIGURAS	ix
	LISTA DE GRÁFICOS	ix
	LISTA DE SIGLAS	x
	RESUMO	xi
	ABSTRACT	xii
1	INTRODUÇÃO	1
1.1	Origem do trabalho	1
1.2	Objetivos do trabalho	3
1.2.1	Objetivo geral	3
1.2.2	Objetivos específicos	3
1.3	Justificativa e importância do trabalho	3
1.4	Metodologia e desenvolvimento da pesquisa	4
1.5	Estrutura do trabalho	5
2	REVISÃO DE LITERATURA	7
2.1	Reconhecimento de Padrões (RP)	7
2.1.1	Extração das características	13
2.1.2	Treinamento	14
2.1.3	Teste	14
2.1.4	Aplicação	15
2.2	Abordagens em reconhecimento de padrões	15
2.2.1	Reconhecimento de padrões estatístico	15
2.2.2	Reconhecimento de padrões sintático	16
2.2.3	Reconhecimento de padrões por redes neuronais	16
2.2.4	Reconhecimento de padrões usando conjuntos difusos	18
2.3	Problemas de reconhecimento de padrões e considerações sobre a vizinhança e conjuntos difusos	18
2.3.1	Visão geral de conjuntos difuso.....	19
2.3.1.1	Conjuntos difusos	20

2.3.1.2	Formação de conjuntos difusos (definição das funções de pertinência)	22
2.3.1.3	Representação de conjuntos difusos	22
2.3.1.4	Representação adicional de vagueza pelo uso de advérbios (<i>hedges</i>)	25
2.3.1.5	Operadores difusos – intersecção, união, complementação....	26
2.4	Atualidades no reconhecimento de padrões.....	28
2.5	FAN – Free Associative Neurons.....	33
3	ESTRUTURA ATUAL DO MODELO FAN COMO MÉTODO PARA RECONHECIMENTO DE PADRÕES	42
3.1	Problema	45
3.2	Notação	45
3.3	Estrutura dos dados de treinamento	46
3.4	Estrutura dos dados de teste	46
3.5	Processo	46
3.6	Treinamento	46
3.6.1	Passo 1: Normalização	47
3.6.2	Passo 2: Determinação do FAN inicializador	47
3.6.3	Passo 3: Construção dos conjuntos difusos para os padrões ..	48
3.6.4	Passo 4: Normalização dos graus de pertinência	49
3.6.5	Passo 5: Determinação da média geométrica (Hedge) do FAN _i e dos graus de pertinência modificados do padrão	50
3.6.6	Passo 6: Determinação da força de representação do padrão	51
3.6.7	Passo 7: Tomada de decisão	52
3.6.8	Passo 8: Penalização e/ou reforço	52
3.6.9	Passo 9: Teste	53
3.7	Visão esquemática do algoritmo de FAN	61
4	APLICAÇÕES DE FAN	63
4.1	Aplicação no reconhecimento de cromossomas do grupo de Denver	64

4.2	Aplicação na resolução do problema do XOR	70
4.3	Mapa FAN	75
4.3.1	A aplicação no estagiamento automático do sono	77
4.4	Utilização de FAN para a construção de um agente para supervisão de alunos em jogos de empresas	79
5	CONCLUSÃO	82
	FONTES BIBLIOGRÁFICAS	85

LISTA DE TABELAS

Tabela 1 – Estrutura dos dados de treinamento	46
Tabela 2 – Estrutura dos dados de teste	46
Tabela 3 – Comparação entre as taxas de erro em diversas aplicações	70
Tabela 4 – Padrões derivados do problema do XOR	70

LISTA DE FIGURAS

Figura 1 – Sistema típico de reconhecimento de padrões.....	12
Figura 2 – Conjuntos difuso e clássico de crianças	20
Figura 3 – Exemplo de função de pertinência triangular	24
Figura 4 – Representação de um número difuso em torno de uma característica	48

LISTA DE GRÁFICOS

Gráfico 1 – Representação do neurônio correspondente à primeira classe em uma aplicação prática após o treinamento.....	54
Gráfico 2 – Representação do neurônio correspondente à segunda classe em uma aplicação prática após o treinamento	55
Gráfico 3 – Representação do neurônio correspondente à terceira classe em uma aplicação prática após o treinamento.....	56
Gráfico 4 – Representação do neurônio correspondente à quarta classe em uma aplicação prática após o treinamento.....	57
Gráfico 5 – Representação do neurônio correspondente à quinta classe em uma aplicação prática após o treinamento.....	58
Gráfico 6 – Representação do neurônio correspondente à sexta classe em uma aplicação prática após o treinamento.....	59
Gráfico 7 – Representação do neurônio correspondente à sétima classe em uma aplicação prática após o treinamento.....	60

Gráfico 8 –	Desempenho (Taxa de Erro x Número de Épocas) nos conjuntos de treinamento e teste na aplicação prática de cromossomas	67
Gráfico 9 –	Desempenho (Taxa de Erro x Número de Épocas) no conjunto de teste na aplicação prática dos cromossomas	68
Gráfico 10 –	Representação de um dos neurônios na aplicação dos cromossomas do grupo de Denver	69
Gráfico 11 –	Representação do primeiro neurônio correspondente à primeira classe no problema de XOR, após o treinamento	73
Gráfico 12 –	Representação do segundo neurônio correspondente à segunda classe no problema de XOR, após o treinamento	74

LISTA DE SIGLAS

FAN – Free Associative Neurons (Neurônios Livre-Associativos)

IA – Inteligência Artificial

RP – Reconhecimento de Padrões

PPD – Processamento Paralelo Distribuído

RESUMO

Nesta tese é apresentado FAN 2002 – *free associative neurons* – como uma abordagem neuro-fuzzy para reconhecimento supervisionado de padrões. Aspectos foram modificados desde a sua publicação inicial, em 1997: a criação de células para a combinação de características é substituída pela inclusão de uma característica extra que representa a combinação das demais; a forma de realizar a penalização em caso de erro é modificada; a inicialização dos neurônios é realizada com valores aleatórios; um *hedge* é usado no teste de um padrão. Os resultados em aplicações são mostrados: no problema de classificação de cromossomos, a comparação com os resultados obtidos por outros métodos apresentados ilustra o desenvolvimento do modelo desde as suas primeiras implementações; A aplicação do modelo em jogos de empresas na construção de um agente; Os conceitos de FAN na construção dos Mapas FAN e a solução do problema do XOR. A apresentação de FAN 2002 visa proporcionar aos interessados na aplicação de soluções em reconhecimento de padrões uma ferramenta útil e de uso geral. A conclusão do trabalho mostra o resumo dos avanços obtidos no desenvolvimento de FAN e aponta para o futuro do método uma possibilidade interdisciplinar de subsidiar pesquisadores de outras áreas a resolverem os seus próprios problemas de classificação e reconhecimento.

Palavras-chave: reconhecimento de padrões; inteligência artificial; FAN; conjuntos difusos.

ABSTRACT

In this dissertation it is presented FAN 2002 – Free Associative Neurons as a neuro-fuzzy approach for supervised pattern recognition. Aspects of FAN were changed since its first approach in 1997: The cells for the features combination are replaced by the inclusion of a single new feature which is composed by the values of the other ones; the way used to perform the penalization is changed; the initialization of the neurons is made through random values; a hedge is included in the test of a pattern. Results obtained in applications are included: on the application of FAN 2002 for the chromosomes recognition is discussed how the method was improved when a comparison of its error tax with the others obtained by other approaches presented is made; concepts of FAN are commented on its application in the construction of the FAN Maps; the application of FAN for the construction of an agent is commented; the solution of the XOR problem. The subject of this work is to support the application of solutions in pattern recognition with a useful and general purpose tool. At the conclusion is summarized the progress of the method and suggested the use of FAN as a way for helping the researches, whose subject is not that of the pattern recognition, to solve their own problems of classification or recognition.

Keywords: Pattern recognition; FAN (Free Associative Neurons); fuzzy sets; Artificial Intelligence.

1 INTRODUÇÃO

1.1 Origem do trabalho

Reconhecimento de Padrões (RP) é uma das áreas de grande abrangência da ciência, estudada em diversos campos de pesquisa como biologia, psicologia, medicina, marketing, finanças, meteorologia, sensoriamento remoto, processamento de imagens, entre outros.

Nos últimos anos, mesmo com o crescente desenvolvimento de técnicas inteligentes, como redes neurais, conjuntos difusos, para o reconhecimento de padrões, ainda permanecem lacunas que são desafios a serem superados.

O objetivo principal na teoria de reconhecimento de padrões é a classificação. Na literatura, existem diversas abordagens para RP, duas destas abordagens, no entanto, estão presentes em praticamente toda a literatura que versa sobre o assunto, podendo ser consideradas como abordagens clássicas de RP. São o RP estatístico e o RP sintático. Outras abordagens mais recentes são a abordagem neural e a abordagem difusa.

Este trabalho tem sua origem na tentativa de integrar técnicas distintas de áreas distintas, mas relacionadas pelo objeto de estudo – classificação de padrões.

Tais áreas – *Inteligência Aplicada, Reconhecimento de Padrões e Teoria da Informação* – apresentam teorias e técnicas complementares; todavia, tentar integrá-las não caracteriza um fim em si, mas um começo para futuras pesquisas.

RP pode ser definido como a ciência que consiste na descrição ou classificação de medidas. Não há dúvidas em relação à importância, utilidade e crescimento no desenvolvimento de tecnologias interdisciplinares que concorrem para a sua realização (SCHALKOFF, 1992).

Em problemas complexos de RP, a abordagem comum utilizada é dividir o problema em dois módulos subsequentes: a extração de características e o módulo de classificação.

Tanto para o primeiro como para o segundo módulo já existem várias técnicas desenvolvidas que vão desde técnicas estatísticas e matemáticas até a de Inteligência Artificial (IA) (BISHOP, 1995), as quais variam dependendo dos tipos de padrões utilizados.

Cada classe deve possuir uma representação independente e autônoma. O treinamento não deve misturar informações de classes distintas. Desta forma, cada classe de interesse no domínio do problema pode ser tratada de forma independente para os propósitos de treinamento e/ou associação com os padrões.

Com base neste princípio, em RAITTZ (1997) é apresentado o modelo FAN (*Free Associative Neurons*) – *Neurônios Livre-Associativos*. Neste trabalho, foi despertada a necessidade de estudos mais aprofundados. Várias novas aplicações foram feitas e as modificações adotadas implicam simplificações que podem tornar o modelo ainda mais genérico.

1.2 Objetivos do trabalho

1.2.1 Objetivo geral

Esta tese tem como objetivo principal apresentar o modelo FAN 2002 como uma abordagem baseada em conjuntos difusos e redes neuronais para o reconhecimento de padrões.

1.2.2 Objetivos específicos

Como objetivos específicos têm-se:

- apresentar o modelo;
- desenvolver um aplicativo para a utilização de FAN em problemas de classificação.
- analisar o desempenho quando aplicado a problemas de classificação de padrões;

1.3 Justificativa e importância do trabalho

Grande parte das técnicas utilizadas computacionalmente para a classificação de padrões consiste na utilização de redes neuronais, por meio de um processo supervisionado de aprendizagem. As informações se encontram como em uma caixa preta (SCHALKOFF, 1992). Isso quer dizer que as informações relativas aos padrões só podem ser acessadas conforme o objetivo específico para o qual foram geradas.

Para a melhor compreensão dos modelos e mesmo para a utilização destas informações conjuntamente com outras técnicas, faz-se necessário o

desenvolvimento de novas estruturas que permitam uma utilização mais ampla das informações resultantes de treinamentos.

A utilização de FAN em RP indica uma possibilidade de obtenção de maior clareza, ainda que não total, o que parece ser um caminho no estudo de novos modelos, quanto à portabilidade das bases de conhecimento e geração de funções de pertinências para a aplicação em conjuntos difusos.

1.4 Metodologia de desenvolvimento da pesquisa

O trabalho foi dividido em etapas as quais são descritas brevemente a seguir.

A primeira etapa constitui o levantamento bibliográfico, onde a partir do qual se levantou informações acerca do estado da arte na área de reconhecimento e classificação de padrões, enfatizando-se questões tais como: o contexto no qual se inserem as aplicações em reconhecimento de padrões; o problema da extração de características em reconhecimento de padrões; técnicas utilizadas para se obter soluções; e as técnicas usadas como suporte para a construção da nova abordagem proposta.

Posteriormente, foram consideradas as possibilidades de adequação da abordagem FAN RAITTZ (1997) relacionada a outras técnicas mais conhecidas, buscando-se obter maior generalidade e clareza na construção, melhores resultados nos experimentos realizados, além de novas aplicações possíveis para o método.

Desenvolveu-se, então, um novo modelo para reconhecimento de padrões, derivado de RAITTZ (1997), o qual sofreu refinamentos sucessivos até se

atingir o estado atual do trabalho. Os citados refinamentos são derivados de uma intensa experimentação e busca de uma sintonia (*tunning*) justa do modelo. Os experimentos e resultados mais interessantes são descritos no capítulo referente às aplicações.

Finalmente, foi desenvolvido um *software*, utilizando-se a linguagem de programação Delphi, para que o uso do método seja viável sem que haja necessidade de conhecimento mais profundo do ponto de vista do usuário, no que se refere à área de reconhecimento de padrões. Uma breve descrição desse produto se encontra descrita no capítulo relativo às aplicações.

1.5 Estrutura do trabalho

A tese é composta de cinco capítulos, seguidos das fontes bibliográficas.

O capítulo 1 - **Introdução** - apresenta aspectos preliminares do trabalho, bem como a apresentação do tema, sua importância, os objetivos do trabalho e sua respectiva estruturação.

No capítulo 2 – **Revisão da literatura** - tem-se uma revisão sobre as principais bases teóricas nas quais foi baseado FAN em seu estágio inicial (RAITZ, 1997), complementada com os aspectos relevantes para desenvolver o estágio atual do modelo apresentado nesta tese.

O capítulo 3 – **Estrutura atual do Modelo FAN como método para reconhecimento de padrões** – apresenta o estado atual do modelo de reconhecimento de padrões usado por FAN, de forma matemática, considerando-se principalmente conceitos de *conjuntos difusos*. Esse capítulo

visa a proporcionar aos interessados na aplicação de FAN subsídios suficientes para a implementação computacional do método.

No capítulo 4 – **Aplicações de FAN** –, encontram-se aplicações que ilustram a relevância do método FAN para a sua aplicação em reconhecimento de padrões, contando resultados em aplicações de interesse acadêmico, bem como comenta implementações baseadas em FAN desenvolvidas por outros autores, com o objetivo de apresentar aplicações práticas do modelo.

O capítulo 5 - **Conclusão** - apresenta as conclusões a respeito do que é mostrado nos outros capítulos e busca mostrar a validade deste trabalho levando-se em conta os seus aspectos de originalidade, não-trivialidade e relevância para a comunidade acadêmica.

2 REVISÃO DA LITERATURA

2.1 Reconhecimento de Padrões (RP)

De acordo com SHROUT (1985), há dois tipos de raciocínio: o indutivo e o dedutivo. Resumidamente, o raciocínio dedutivo ocorre quando se parte de uma generalização, de uma teoria ou premissa e raciocinamos em direção ao específico. O raciocínio indutivo acontece quando se parte de dados individuais e daí raciocinamos para uma espécie de teoria.

Continua SHROUT: ao ver uma maçã cair da árvore, Newton percebeu um fenômeno isolado e começou a raciocinar sobre ele para criar sua teoria geral. Com efeito, Newton raciocinou indutivamente sobre tudo o que cai. Então, reuniu mais alguns dados e perguntou-se: – O que isso demonstra? – Assim, propôs a teoria da força da gravidade.

O mesmo autor conclui que o raciocínio indutivo implica reunir dados e induzir uma teoria capaz de esclarecer os fatos.

RUIZ (1986) concorda com SHROUT afirmando que a indução é um processo de raciocínio inverso ao processo dedutivo. Enquanto a dedução parte de enunciados mais gerais para se chegar a uma conclusão menos geral, a indução caminha do registro de fatos singulares ou menos gerais para se chegar a uma conclusão desdobrada ou ampliada relativa a um enunciado mais geral.

No dicionário de psicologia DORON & PAROT (1991), encontra-se o seguinte texto referente ao termo padrão e reconhecimento de padrões:

“O termo inglês *pattern* pode ocorrer em contextos muito diversos, e ser traduzido, conforme o caso, por padrão, configuração, estrutura, boa forma,

modelo (no sentido de modelo a ser reproduzido), tipo, procedimento geral. Em algumas de suas acepções inglesas, ele se reveste de um valor semântico que parece não ter equivalência em nenhuma palavra do português, o que explica, talvez, sua adoção na linguagem científica como *pattern*. A esse respeito, é particularmente ilustrativo o sentido de *pattern* em *pattern recognition*, ou reconhecimento de *pattern*: trata-se da capacidade de percepção (capacidade de que hoje se procura dotar as máquinas) para identificar uma mesma forma, por exemplo, de uma letra ou de um algarismo por entre suas múltiplas realizações nas escritas individuais. A palavra ‘forma’ que realmente tem em português, entre outros, esse sentido, não parece evocar, no mesmo grau, a dupla conotação de flexibilidade e de estrutura. Casos desse tipo, que propiciam o empréstimo lexical direto, favorecem, ao mesmo tempo, sua generalização em acepções para as quais o português dispõe, no entanto, de um equivalente”.

A partir dos parágrafos anteriores, pode-se forjar a definição para o termo *Reconhecimento de Padrões* (RP) utilizada ao longo desse trabalho. Nesse sentido, reconhecimento de padrões é a aplicação de técnicas de raciocínio indutivo sobre um conjunto de dados conhecidos com o objetivo de construir um modelo conforme o qual novos dados não conhecidos possam ser classificados de forma conveniente.

Sob essa perspectiva, não é adequado afirmar que o *Reconhecimento de Padrões* ou *pattern recognition* seja uma área necessariamente nova do conhecimento humano.

Desde a antigüidade, tem-se buscado a associação dos fatos a modelos gerais. Se por um lado, a forma de definir e tratar os padrões nas culturas antigas serviu para o desenvolvimento de superstições e mal-entendidos, por outro, esse é o único elo que torna possível à civilização contemporânea conhecer essas culturas.

Alguns tradicionalistas procuram ainda hoje respostas para questões atuais a partir desses modelos antigos.

Conforme MEBES (1998),

“De acordo com a tradição, os sacerdotes de Menphis, prevendo a queda da civilização egípcia, ocultaram seus conhecimentos sob a forma de um baralho que, hoje em dia, é conhecido sob o nome de Tarô e o legaram aos profanos, sabendo que, devido ao hábito do jogo, tais conhecimentos serviriam à posteridade.”

Nessa citação, o autor confere ao baralho do Tarô uma forma de transmissão de conhecimento, e remonta a sua origem à remota antigüidade dos templos egípcios.

Outros autores, como D'ALVEYDRE (1997), atribuem os padrões de construções das antigas catedrais européias a formas muito sistemáticas, oriundas do pensamento da ancestralidade humana. Mesmo não sendo considerado o valor estritamente científico dos padrões registrados em tais monumentos, é inegável sua grandiosidade nos campos da arte e arquitetura.

Outras formas de construção de modelos de padrões, baseados em símbolos, podem ser encontradas ao longo da história em todas as artes divinatórias, no I-Ching, na astrologia etc.

Mesmo na linguagem empregada na comunicação do dia-a-dia, nota-se a padronização de símbolos – das palavras – que são adotados para relacionar fatos e idéias a eles subjacentes. Então, pode-se concluir que existe uma relação entre a formação de modelos para a representação de padrões e a comunicabilidade de idéias; ou seja, o estabelecimento de relações que agrupam fatos. Essa é a demanda a qual as técnicas de RP visam a atender.

Neste caso, os dados relativos aos padrões são expressos em valores matemáticos e as técnicas conhecidas e utilizadas para RP são ferramentas que buscam integrar os recursos matemáticos e computacionais com o objetivo de encontrar soluções para problemas.

Entende-se que, no domínio da inteligência artificial, encontram-se os melhores recursos para o tratamento de problemas com as características de reconhecimento de padrões. Em SCHALKOFF (1992), observa-se uma relação entre inteligência artificial e reconhecimento de padrões, que pode ser resumida no parágrafo seguinte.

A inteligência computacional é uma das tecnologias mais emergentes nos últimos anos. Um dos componentes freqüentemente necessários no desenvolvimento de sistemas inteligentes é o reconhecimento de padrões.

Em linhas gerais, *Reconhecimento de Padrões* (RP) é a ciência que trata da descrição ou classificação (reconhecimento) de medidas. Não há dúvidas de que o RP constitua uma tecnologia útil e importante que vem se desenvolvendo rapidamente com interesses e participações interdisciplinares. Assim, RP não se reduz a uma forma de aplicação, mas apresenta um leque de abordagens possíveis que podem ou não ser relacionadas entre si.

De fato, o reconhecimento de padrões envolve uma gama enorme de problemas de processamento de informação, com grande significância prática, desde o reconhecimento de voz e de caracteres feitos manualmente, até à detecção de erros em equipamentos ou diagnósticos médicos. Não é possível comentar todas as áreas em que seria útil e, por vezes, mesmo necessário, aplicar RP. Tal comentário se justifica se for levado em conta que as pessoas

aplicam seus próprios métodos de reconhecimento de padrões em praticamente todas as áreas da atividade humana.

Isso porque a racionalidade pura não é suficiente para o bom desempenho na resolução da maioria dos problemas. Tome-se como exemplo o jogador de xadrez. A análise de todas as possibilidades em uma dada posição complexa é virtualmente impossível para um jogador que, no entanto, se for hábil, desenvolve técnicas próprias para identificar os padrões das várias posições as quais se deparou em suas experiências e usar estas técnicas com o fim de obter posições vantajosas em seus jogos.

Na prática computacional, pode-se enumerar algumas áreas em que vem sendo usado o RP:

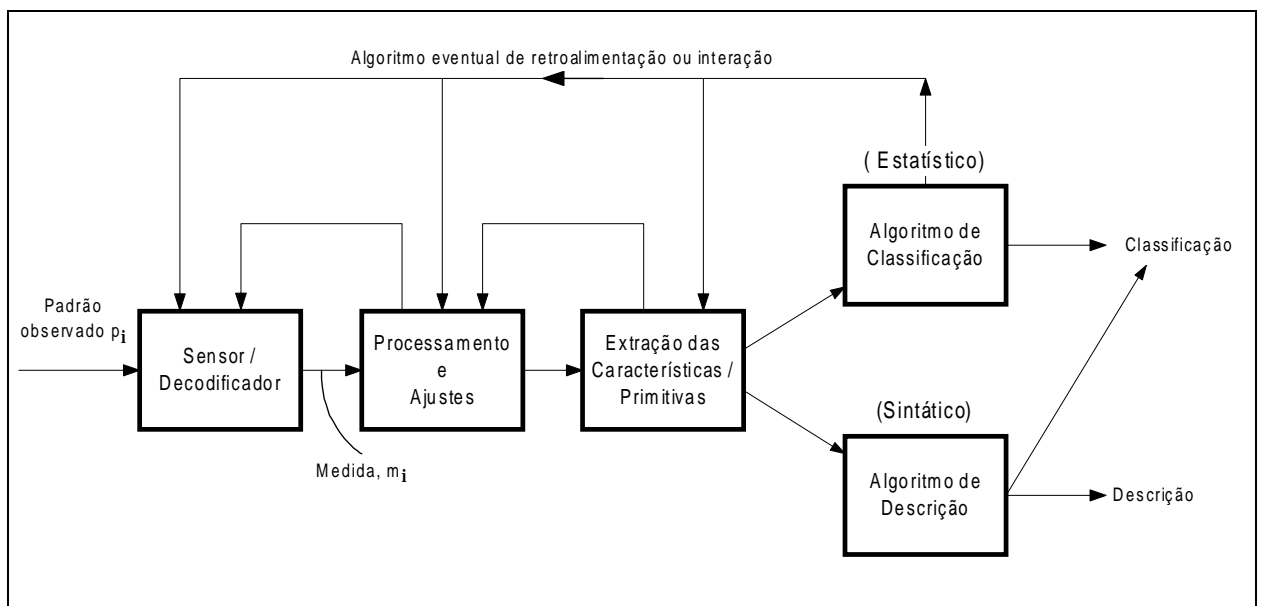
- processamento de imagens;
- visão computacional;
- classificação/análise dos sinais de radar;
- reconhecimento de fisionomia;
- reconhecimento de voz;
- tratamento de impressões digitais;
- reconhecimento de caracteres;
- análise de manuscritos;
- análise de eletrocardiogramas e eletroencefalogramas;
- diagnósticos médicos.

ROSS (1995) define *Reconhecimento de Padrões* como um processo de identificação de estruturas nos dados por comparações com estruturas conhecidas.

Por sua vez, KLIR (1995) entende que o reconhecimento de padrões pode ser entendido como um processo pelo qual se buscam estruturas nos dados e classificam-se essas estruturas de acordo com categorias tais que o grau de associação é maior entre as estruturas da mesma categoria e menor entre as categorias de estruturas diferentes. As categorias relevantes são usualmente caracterizadas por estruturas prototípicas derivadas da experiência passada. Cada categoria pode ser caracterizada por mais de uma estrutura prototípica.

A seguir, apresentar-se-á uma estrutura típica de um sistema de reconhecimento de padrões, conforme sugerido na *Figura 1*:

Figura 1 – Sistema típico de reconhecimento de padrões



Fonte: Schalkoff, 1992.

Tal estrutura consiste em um sensor – uma câmera, por exemplo –, um mecanismo de extração de características – algoritmo –, e um algoritmo de descrição ou classificação, a depender da aplicação. Complementarmente, é usual que alguns dados que já tenham sido classificados ou descritos estejam disponíveis para treinar o sistema.

Esse conjunto de dados é chamado de *conjunto de treinamento*. A validação da eficácia na utilização de um sistema de RP é testada em um conjunto de dados não usado para o treinamento.

Pode-se, de forma mais simples, definir-se um sistema de RP em outros quatro aspectos ou fases que são comentados a seguir.

2.1.1 Extração das características

É o processo pelo qual um problema é simplificado de forma a ser possível o seu tratamento computacional. Na prática, o volume de informações associado a um padrão é muito grande, tendendo, não raro, à infinidade. Resumi-las, portanto, torna-se uma questão essencial. O processo é feito através de cálculos de médias, desvios-padrões e escolha das características mais relevantes por meio de transformações matemáticas em geral. Essas são as formas mais utilizadas.

Em SOUZA (1999), pode-se encontrar uma forma bastante interessante de redução de dimensionalidade pela geração de índices recursivos. Ocorre freqüentemente no senso comum um equívoco relacionado à extração de características. Isso porque há uma tendência de se supor que, se for utilizado o maior número possível de informações disponíveis em um dado problema, obter-se-ão melhores soluções, fato que raramente ocorre.

O abuso no volume de dados implica sobrecarga dos sistemas tornando-os demasiadamente restritivos. Por exemplo: se uma pessoa tomasse a cada momento que observasse um objeto todas as informações que tem disponíveis a respeito dele, jamais poderia reconhecê-lo, porque tal objeto se transforma a

cada instante, quer seja pela sua mudança de estado ou pela mudança no plano de observação adotado.

Por outro lado, a falta de informações leva à classificação falsa, uma vez que se pode considerar objetos semelhantes de classes diferentes como sendo o mesmo objeto. Para ocorrer uma boa extração de características, deve-se procurar encontrar o conjunto resumido de características de forma adequada ao problema em tratamento.

2.1.2 Treinamento

Uma vez extraídas as características, obtém-se delas um conjunto, que deve representar os padrões do modelo. Nesse ponto, é suposto que se tenha em mãos um conjunto de padrões já classificados, o qual pode ser apresentado ao sistema para se construir o modelo de reconhecimento de padrões. Diz-se normalmente que nessa fase ocorre o aprendizado.

Quando as classes às quais os padrões de treinamento pertencem são conhecidas, os sistemas de RP são definidos como supervisionados. Vale esclarecer que ao longo desse trabalho só serão tratados os métodos supervisionados.

2.1.3 Teste

Uma vez treinado o sistema, sua capacidade de classificação é avaliada em um algoritmo de teste, feito sob padrões do mesmo tipo daqueles utilizados no treinamento, cujas classes também são conhecidas, sendo que eles não tenham sido usados para o treinamento.

Por assim dizer, o sistema *não os conhece*. A taxa de acerto obtida sobre o conjunto de testes indica a performance do sistema.

2.1.4 Aplicação

Tendo sido o sistema treinado e testado, ele pode ser utilizado para reconhecer padrões. Isto significa que padrões não supervisionados podem ser relacionados às classes do domínio do problema uma vez que sejam conhecidas apenas suas características. O percentual de acertos esperado na classificação deve ser o mesmo que ocorre na fase de testes.

2.2 Abordagens em reconhecimento de padrões

A seguir, são comentadas algumas abordagens usadas para o tratamento de problemas de reconhecimento de padrões.

2.2.1 Reconhecimento de padrões estatístico

De acordo com BISCHOP (1995), a forma mais geral e natural de formular soluções para o reconhecer padrões é o RP estatístico, por meio do qual é reconhecida a natureza estatística tanto da informação que se quer representar quanto dos resultados que devem ser expressos.

O reconhecimento de padrões estatístico é uma área de conhecimento bem estabelecida, que possui uma longa história. SCHALKOFF (1992) preconiza que o reconhecimento de padrões estatístico, como sugere o próprio nome, assume uma base estatística para os algoritmos de classificação.

Um conjunto de medidas, que denotam as características, é extraído dos dados de entrada e usado para associar cada vetor de características a uma de entre c classes.

Presume-se que as características são geradas por um estado da natureza, e que existe um modelo subordinado a um estado com um conjunto de probabilidades e/ou funções de densidade de probabilidades correspondente passível de ser usado para representá-lo.

2.2.2 Reconhecimento de padrões sintático

Muitas vezes, as informações significativas em um padrão não constituem apenas presença ou ausência de valores numéricos de um conjunto de características. Ao invés disto, a inter-relação e a interconexão das características produzem informações estruturais importantes que facilitam a descrição ou classificação. Esta é a base do RP sintático. Portanto, torna-se necessário, ao se usar a abordagem de RP sintático, quantificar e extrair as informações estruturais importantes para poder avaliar a similaridade estrutural entre os padrões.

Tipicamente, o RP sintático formula uma descrição hierárquica de padrões complexos, construída a partir de outros subpadrões mais simples, sendo que no nível mais baixo se encontram os elementos mais simples, extraídos dos dados de entrada que são chamados de primitivas.

2.2.3 Reconhecimento de padrões por redes neurais

Alguns trabalhos, como é o caso de BISCHOP (1995), adotam o ponto de vista de que a abordagem de RP, por meio do uso de redes neurais, é uma

extensão do RP estatístico convencional, sendo a segunda forma considerada um aperfeiçoamento da primeira.

Uma rede neuronal ou connexionista é definida em CICHOKI & UNBEHAUEN (1993) como um sistema de processamento de sinal ou informação, composto por um grande número de elementos simples de processamento, chamados neurônios artificiais, ou simplesmente nós, que são interconectados por elos diretos chamados de conexões, que colaboram para realizar um processamento paralelo distribuído (PPD) e uma tarefa computacional desejada.

A tarefa computacional citada no parágrafo anterior pode ser: dado um vetor de características, deve-se associá-lo a uma classe em um determinado problema (i. e. reconhecimento de padrões).

Na maioria das aplicações de RP, usando-se redes neurais, são estabelecidas conexões entre os valores das características que definem os padrões, com as camadas intermediárias de neurônios e com todas as classes objetivo, consideradas a camada de saída.

O treinamento da rede é realizado corrigindo os pesos nas conexões para estabelecer as relações entre as características e as classes para prover a melhor discriminação possível entre os padrões de classe diferentes.

Assim, ao ser apresentado à rede um novo padrão, ela indicará a classe que o represente melhor na camada de saída. É importante salientar que o conceito de neurônio usado nas redes connexionistas é diferente do atribuído às unidades FAN; este conceito será abordado posteriormente.

Existem vários trabalhos publicados que tratam exclusivamente do reconhecimento de padrões utilizando-se as redes neuronais. O de BISHOP (1995) é um deles.

Por sua vez, DANDOLINI (1995) também faz um estudo detalhado de diversos tipos de redes neuronais aplicadas para reconhecer padrões.

Ainda, TODESCO (1995) obteve resultados em reconhecimento de grupos de cromossomas pelo uso de uma rede neuronal conhecida como *Radial Basis Function* (RBF). Alguns resultados das comparações utilizadas neste trabalho se encontram nessas publicações.

2.2.4 Reconhecimento de padrões utilizando conjuntos difusos

A utilidade da teoria dos conjuntos difusos no reconhecimento de padrões já foi reconhecida em meados da década de 1960, e hoje a literatura a respeito de RP difuso é bastante extensa, conforme demonstra KLIR (1995). Uma revisão dos aspectos da teoria dos conjuntos difusos mais relacionados a esse trabalho é apresentada no item seguinte.

2.3 Problemas de reconhecimento de padrões e considerações sobre a vizinhança e conjuntos difusos

A linha básica utilizada para a construção deste tópico é a adotada por DURKIN (1994), reforçada por ROSS (1995), KLIR (1995) e KOSKO (1992).

2.3.1 Visão geral de conjuntos difusos

O primeiro a utilizar conceitos de lógica difusa foi LUKASIEVICZ na década de 1920, ao estudar a representação matemática de termos como *alto*, *velho* ou *quente*. Tais termos não podem ser adequadamente representados pela lógica aristotélica, baseada na verdade ou falsidade absoluta, 0 ou 1.

O matemático polonês desenvolveu um sistema de lógica, em que os graus de verdade possíveis se estendem aos números reais no intervalo $[0,1]$. LUKASIEVICZ utilizou um número nesse intervalo para representar a *possibilidade* de uma declaração dada ser verdadeira ou falsa. Por exemplo, a possibilidade de uma pessoa de 1,80 m de altura ser realmente alta pode ser relacionada ao valor 0,9; é muito provável que tal pessoa seja alta. Tal raciocínio levou ao desenvolvimento de uma técnica formal para tratar o raciocínio inexato, denominado teoria da possibilidade.

Em 1965, ZADEH estendeu o trabalho a um sistema formal de lógica matemática. O trabalho de ZADEH possui uma relevância especial, por tê-lo transformado em uma ferramenta que permite o uso prático nos meios técnicos e científicos de conceitos antes essencialmente filosóficos.

Essa *ferramenta* valiosa para o tratamento de termos da linguagem natural é chamada *lógica difusa*. Define-se a lógica difusa como o ramo da lógica que usa graus de pertinência em conjuntos, em vez da pertinência e ausência absolutas: verdadeiro ou falso.

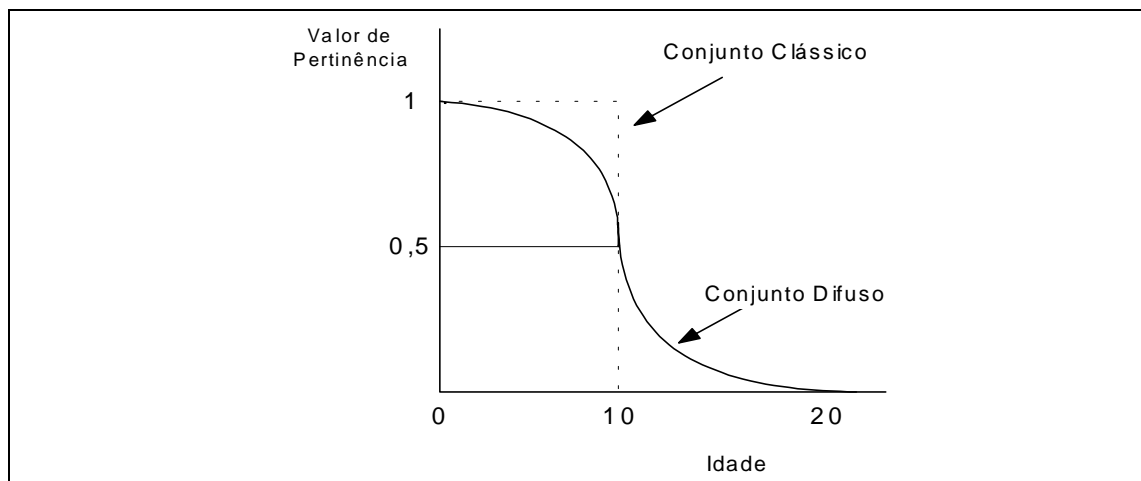
2.3.1.1 Conjuntos difusos

A teoria tradicional dos conjuntos olha para o mundo como se ele fosse preto e branco. Isso significa que um objeto pertence ou não pertence a um determinado conjunto. Se for considerado, por exemplo, o conjunto das crianças, a teoria clássica dos conjuntos usa a definição de um limite rígido de idade a partir da qual as pessoas deixam de ser crianças, sendo que todos os indivíduos com idade aquém deste limite são membros do conjunto com valor 1 e os primeiros o são com valor 0.

Então, se o limite de idade for estabelecido em dez anos, alguém que complete onze deixa de ser a criança que fora no momento anterior.

A lógica difusa é mais razoável na interpretação de casos como o exposto acima. No uso de conjuntos difusos com graus de pertinência entre 0 e 1 para os seus membros, a associação membro/conjunto é mais natural. Por exemplo, uma pessoa com 5 anos pode ser associada ao conjunto das crianças com o valor de pertinência 0,9; já uma de 13 anos, com o valor 0,1.

Figura 2 – Conjuntos difuso e clássico de crianças



Fonte: John Durkin, 1994

Os conjuntos difusos proporcionam uma transição suave dos limites ilustrados na *Figura 2*. O eixo-x, definido como universo de discurso, representa as idades das pessoas. O eixo-y é o dos valores de pertinência ao conjunto difuso. Este conjunto, das crianças, mapeia as idades em relação aos seus valores de pertinência. Observando a figura, uma pessoa de 11 anos não deixa de ser criança subitamente. A pessoa vai deixando de pertencer ao conjunto das crianças, de modo gradual, conforme vai assimilando as características próprias do mundo adulto, com o aumento da idade.

Assim, a definição formal de conjunto difuso é a seguinte:

Seja X o universo de discurso, com os elementos de X denotados por x . Um conjunto difuso A de X é caracterizado por uma função de pertinência $\mu_A(x)$ que associa cada elemento x com um grau de pertinência em A .

Em contraste com a teoria da probabilidade, que atribui probabilidades a um dado evento baseadas na sua frequência de ocorrência, a lógica difusa atribui valores aos eventos baseados na função de pertinência definida como:

$$\mu_A(x): X \rightarrow [0,1]$$

Na teoria dos conjuntos difusos o evento ou elemento x é associado a uma função de pertinência μ . Esse valor representa o grau ao qual o elemento x pertence ao conjunto difuso A .

$$\mu_A(x) = \text{Grau}(x \in A)$$

O valor de pertinência de x é delimitado pela seguinte relação:

$$0 \leq \mu_A(x) \leq 1$$

Os conjuntos difusos são uma extensão da teoria tradicional dos conjuntos, que generaliza o conceito de pertinência, usando a função de pertinência μ que

retorna um valor entre 0 e 1, representando o grau de pertinência que um objeto x tem em relação ao conjunto A .

2.3.1.2 Formação de conjuntos difusos (definição das funções de pertinência)

Para se representar conjuntos difusos em computadores, é necessário definir suas funções de pertinência. Para fazê-lo, é conveniente consultar pessoas que conheçam bem o domínio do problema a ser tratado, ou que cuja opinião sobre a significação da variável a ser modelada seja significativa para sua boa representação.

Numerosos métodos para construção de funções de pertinência se encontram descritos na literatura dos conjuntos difusos e quase sempre são baseados em julgamentos de especialistas no assunto que se quer representar.

Tais métodos estão classificados em duas categorias fundamentais: diretos e indiretos. Nos métodos diretos, espera-se que o especialista responda a questões de vários tipos que definam explicitamente as funções de pertinência. Nos métodos indiretos, as informações fornecidas pelos especialistas são mais genéricas e afetam as funções construídas apenas implicitamente (KLIR, 1995).

2.3.1.3 Representação de conjuntos difusos

A forma de representação dos conjuntos difusos utilizada pela literatura possui algumas características que podem confundir o iniciante no estudo da teoria. Segue-se uma introdução sobre a representação formal dos conjuntos difusos.

Dado um universo de discurso X e um conjunto difuso A definido nele, e um conjunto de elementos discretos de n elementos $\{x_1, x_2, \dots, x_n\}$. O conjunto difuso A define a função de pertinência $\mu_A(x)$, que mapeia os elementos x_i de X , em relação aos graus de pertinência em $[0, 1]$.

Os graus de pertinência indicam qual é o grau que x_i pertence a A . Para um conjunto discreto de elementos, uma representação = conveniente é feita usando um vetor:

$$A = (a_1, a_2, \dots, a_n)$$

onde

$$a_i = \mu_A(x_i)$$

Para uma representação mais clara, o vetor normalmente inclui o símbolo “/” que associa o valor de pertinência a_i com sua coordenada x_i :

$$A = (a_1/x_1, a_2/x_2, \dots, a_n/x_n)$$

É importante observar que o sinal “/” não representa fração, mas a relação entre valor e pertinência. A notação padrão dos conjuntos difusos representa a união das dimensões dos vetores como é mostrado na sequência; em que “+” representa a notação de união Boleana:

$$A = \mu_1/x_1 + \mu_2/x_2 + \dots + \mu_n/x_n$$

ou

$$A = \sum_{i=1}^n \mu_i/x_i$$

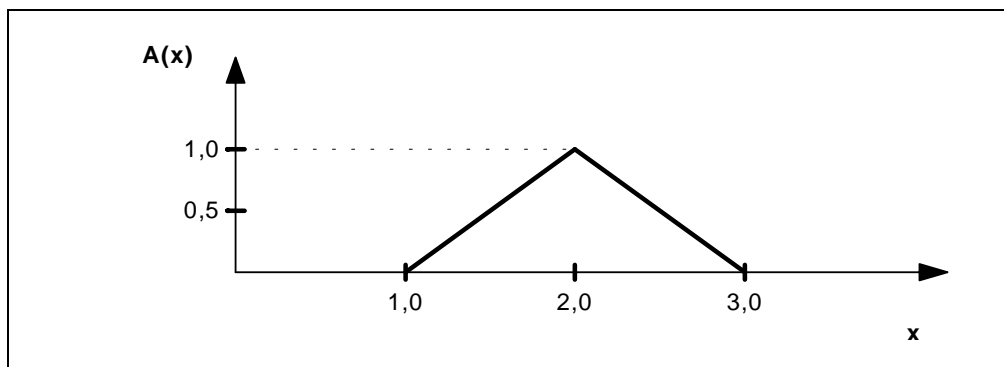
Se X é uma função contínua, então o conjunto A pode ser representado da seguinte forma:

$$A = \int_x \mu_A(x_i) / x_i$$

Para um conjunto contínuo de elementos, é necessário usar funções para mapear os elementos em relação aos seus graus de pertinência.

Algumas funções típicas são as *sigmoidais*, *gaussianas* e *pi*, que são curvas suaves e freqüentemente representam bem os dados. Porém, sob o ponto de vista computacional, nem sempre são as mais convenientes. Se a economia no tempo de processamento é mais relevante em uma aplicação que o critério da suavidade, pode-se usar funções mais simples e em muitos casos eficientes como funções triangulares:

Figura 3 – Exemplo de função de pertinência triangular



Fonte: Raittz, 1997

$$A(x) = \begin{cases} p(x-r) + 1 & \text{quando } x \in [r - 1/p, r], \\ p(r-x) + 1 & \text{quando } x \in [r, r + 1/p], \\ 0 & \text{para outros casos.} \end{cases}$$

Onde r é o valor para o qual se deseja que o grau de pertinência seja 1 ($r=2$, no exemplo) e p é o parâmetro que determina a taxa com a qual, para cada x , a função cresce ou decresce, com o aumento da diferença $|r-x|$.

2.3.1.4 Representação adicional de vagueza pelo uso de advérbios

(*hedges*)

No diálogo, é comum se utilizar elementos enfáticos para reforçar os significados de declarações feitas. Para alcançar maior grau de expressividade numa interlocução, utilizam-se os advérbios: Fulano é *muito* alto, Sicrano é *um pouco* alto. A teoria dos conjuntos difusos possui recursos que podem ser usados para captar a ênfase provocada pelo uso de advérbios. São os *hedges*.

Um *hedge* modifica matematicamente um conjunto difuso já existente, de forma a incorporar ao seu significado o efeito provocado pelo uso de um advérbio.

Para ilustrar este efeito, tem-se como exemplo uma pessoa de 1,80 m que pode ser considerada uma pessoa alta em um conjunto de pessoas altas com 0,7 de pertinência. Entretanto, este mesmo indivíduo pode ser considerado muito alto, com pertinência 0,49, se for aplicado um *hedge* de concentração (muito) sobre sua situação de alto neste caso:

$$\mu_{\text{con}(A)}(x) = (\mu_A(x))^2$$

A aplicação da potência $n=2$, no exemplo, diminui o valor da pertinência dos elementos que têm valores de pertinência pequenos e aumenta o rigor do conceito expresso pelo conjunto difuso original. Existem outros tipos de *hedges*. Como o de dilatação:

$$\mu_{\text{dilat}(A)}(x) = (\mu_A(x))^{0,5}$$

que diminui a restrição do conceito. E outros conforme DURKIN (1994).

2.3.1.5 Operadores difusos

Intersecção (t-norma)

Na teoria clássica dos conjuntos, a intersecção entre dois conjuntos contém os elementos comuns a ambos. Na teoria dos conjuntos difusos, entretanto, o elemento pode pertencer parcialmente aos dois conjuntos, ainda que não pertença completamente a nenhum deles.

Assim, quando é considerada a intersecção desses conjuntos, não se pode dizer que um elemento possa pertencer mais ao conjunto da intersecção do que a qualquer um dos conjuntos originais. De acordo com tal premissa, o operador difuso padrão usado para gerar a intersecção entre dois conjuntos difusos A e B definidos em X é dada por:

$$\begin{aligned}\mu_{A \wedge B}(X) &= \min(\mu_A(x), \mu_B(x)) \quad \text{para todo } x \in X \\ &= \mu_A(x) \wedge \mu_B(x) \\ &= \mu_A(x) \cap \mu_B(x)\end{aligned}$$

O símbolo \wedge ("E" lógico) é usado para representar o operador "min" que simplesmente toma o mínimo entre os valores em consideração. Exemplo:

$$\text{ALTO} = (0/5, 0,2/5,5, 0,5/6, 0,8/6,5, 1/7)$$

$$\text{BAIXO} = (1/5, 0,8/5,5, 0,5/6, 0,2/6,5, 0/7)$$

$$\mu_{\text{ALTO} \wedge \text{BAIXO}}(x) = (0/5, 0,2/5,5, 0,5/6, 0,2/6,5, 0/7)$$

A intersecção entre os conjuntos dados por ALTO e BAIXO, dada por $\mu_{\text{ALTO} \wedge \text{BAIXO}}(x)$, pode ser entendida como as pessoas de média estatura, por exemplo, e usada como um novo conjunto difuso.

União (t-conorma)

Outra forma de combinar conjuntos difusos é por meio de sua união. A união de dois conjuntos é compreendida como o conjunto dos elementos que pertencem a pelo menos um deles (ou a ambos). Assim, os elementos do conjunto união não podem possuir valor de pertinência menor que o que possuía em qualquer um dos conjuntos originais. A forma padrão que a lógica difusa usa para obter a intersecção entre dois conjuntos é a seguinte:

$$\begin{aligned}\mu_{A \vee B}(x) &= \max(\mu_A(x), \mu_B(x)) \quad \text{para todo } x \in X \\ &= \mu_A(x) \vee \mu_B(x) \\ &= \mu_A(x) \cup \mu_B(x)\end{aligned}$$

O símbolo \vee (chamado “ou” lógico) é usado na teoria dos conjuntos difusos para representar a operação “max”, que toma o valor máximo dentre os valores em consideração. Exemplo:

$$\text{ALTO} = (0/5, 0,2/5,5, 0,5/6, 0,8/6,5, 1/7)$$

$$\text{BAIXO} = (1/5, 0,8/5,5, 0,5/6, 0,2/6,5, 0/7)$$

$$\mu_{\text{ALTO} \vee \text{BAIXO}}(x) = (1/5, 0,8/5,5, 0,5/6, 0,8/6,5, 1/7)$$

A união entre os conjuntos dados por ALTO e BAIXO, dada por $\mu_{\text{ALTO} \vee \text{BAIXO}}(x)$, pode ser entendida como as pessoas que não são de estatura média, por exemplo, e também usada como um novo conjunto difuso.

Complementação

Dado um conjunto difuso A , é possível encontrar o seu complemento $\sim A$ por meio da seguinte expressão:

$$\mu_{\sim A}(x) = 1 - \mu_A(x)$$

Por exemplo, dado o conjunto das *peçoas altas* usado nos exemplos anteriores, este operador pode ser usado para criar o conjunto das *peçoas não altas* ou *peçoas medianas* ou *baixas*.

$$\mu_A(x) = \text{ALTO} = (0/5, 0,2/5,5, 0,5/6, 0,8/6,5, 1/7)$$

$$\mu_{\sim A}(x) = \text{NÃO ALTO} = (1/5, 0,8/5,5, 0,5/6, 0,2/6,5, 0/7)$$

2.4 Atualidades no reconhecimento de padrões

Durante os últimos 20 anos, consideráveis desenvolvimentos foram obtidos em processos de monitoramento on-line e in situ. Variáveis não mensuráveis foram monitoradas por estimadores observador/estado e sensores por software. SCHÜGERL (2001) faz uma revisão desses recentes desenvolvimentos em engenharia de bioprocessos, discutindo a avaliação da morfologia e psicologia *mycelial* por imagem digital. Inclui monitoramento do processo de formação de produtos pela análise da injeção de fluxo, vários tipos de métodos cromatográficos e espectroscópicos como também por biosensores. Trabalha com controle avançado de variáveis de avaliação indireta de processos por meio de estados de estimação/observador, com o uso de modelos híbridos e estruturados, sistemas especialistas e reconhecimento de padrões por processos de otimização.

Modelos computacionais para aprendizado de palavras cuja aquisição é realizada diretamente por entradas sensoriais multimodais de ruídos têm sido implementados por ROY e PENTLAND (2002) em um sistema usando novas formas de processamento de voz, visão computacional e algoritmos de aprendizado de máquina. Os autores demonstram a possibilidade do uso de

técnicas de reconhecimento de padrões através de sensores e aprendizado de máquina para implementar modelos cognitivos que podem processar dados com ruídos sem a necessidade da transcrição e anotações humanas. O modelo é construído em cima de um conjunto de suposições feitas para implementar o modelo computacional.

AURA, é uma nova tecnologia de busca e associação de padrões baseada em redes neurais desenvolvida por AUSTIN (2000). A tecnologia explora o treinamento simples e rápido de memórias da matriz de correlações e sua habilidade em associar ruídos e dados incompletos e tem uma efetiva implementação de hardware. Esta aquisição é feita pelo armazenamento de dados de características elementares e sua aplicação é útil em uma gama muito grande de problemas. O AURA supera as severas limitações das arquiteturas das redes neurais padrões nas aplicações de armazenagem de dados, incluindo o problema de escalabilidade. A derivação de uma arquitetura modular permite ao sistema lidar com variações nos elementos de entrada.

A combinação de técnicas como as redes neurais, o reconhecimento de padrões e aspectos da lógica difusa é empregada por GROHMAN e DHAWAN (2001) no diagnóstico de câncer de mama. A capacidade de generalização dos modelos de reconhecimento de padrões é discutida, bem como a sua capacidade em permitir a compreensão do espaço das características envolvidas. A qualidade das aplicações que utilizam redes neurais com *feedforward backpropagation* é questionada por esses autores em relação aos referidos critérios. Um modelo *neuro-fuzzy* construído baseado em funções de pertinências convexas obteve performance com dois tipos de função base:

sigmoidais e funções em forma de sino. O principal resultado do trabalho é a clareza nas informações resultantes no treinamento.

Para minimizar a utilização de herbicidas é usada a rede neuronal SOM (*self organizing map*). A classificação busca separar o material útil da colheita daquele constituído por ervas daninhas pelo uso de propriedades espectrais. O método proposto por MOSHOU et al. (2001) é comparado com um classificador Bayesiano apresentado como uma rede neuronal probabilística e mostra vantagens em relação a ele. O modelo também é comparado com o método *backpropagation* apresentando convergência mais rápida.

Várias redes neurais são usadas conjuntamente por ZHOU (2002) para aumentar o grau de confiança na detecção de câncer de pulmão através da análise de imagens de células. O método chamado NED (*Neural ensemble-based detection*) é dividido em duas etapas. A primeira detecta se a célula é normal ou cancerígena. A segunda é usada para definir qual é o tipo de câncer ao qual a célula anormal é relacionada ou mesmo se a célula é normal (corrigindo o problema de classificação falsa). Os autores justificam a importância de seu trabalho comentando que ainda não há um caminho rigorosamente definido mostrando como obter sucesso na utilização de redes neurais no problema específico abordado por eles.

Ao abordar-se a visão computacional de objetos depara-se com a dificuldade que a maior parte dos algoritmos disponíveis neste sentido são limitados a apenas um dos aspectos relacionados à visão: detecção, reconhecimento, alinhamento, estimação de pose, análise de cenas, movimento, etc. Um algoritmo que incorpora a diversidade destes aspectos é o LGM (*Labeled graph Matching*) e conforme SHAMS et al. (2001) é o melhor

disponível para a finalidade. Mesmo assim os autores apresentam uma nova abordagem para o LGM batizada de LGM+, onde vantagens em relação ao modelo original são obtidas.

A rede neuro-fuzzy Falcon-mart é como no caso anterior uma derivação de um modelo já publicado chamado Falcon-art. Em QUEK & TUNG (2001) estão listados os motivos da nova implementação que têm por finalidade resolver os defeitos da original que são: a baixa performance da rede quando as classes de entrada são muito similares; a baixa tolerância a ruídos nos padrões de treinamento; a dependência de um parâmetro de erro para o término do treinamento; a queda na eficiência do aprendizado resultante do uso de dados de treinamento complementares. O desempenho de Falcon-mart é avaliado por meio de experimentos. Em um dos testes é usado o conjunto da Iris de Fisher, FISHER (1936). Comparações de desempenho da rede são realizadas em relação ao *multi-layered perceptron* (MLP) e o classificador *k-nearest-neighbor* (KNN), ambas implementações clássicas. A forma de trabalho da rede foca a formação de conjuntos difusos a partir dos dados de treinamento e formação de regras difusas para poder representar com qualidade o domínio dos problemas abordados.

Caracteres Kanji são ideogramas de origem chinesa, que possuem características interessantes para testar métodos de reconhecimento de padrões, especialmente no caso em que eles são manuscritos. As dificuldades em reconhecê-los estão relacionadas aos seguintes aspectos: o grande número de categorias de caracteres; a complexidade inerente às formas que os constituem; a existência de caracteres semelhantes e a grande variedade de deformação derivada da escrita manual. Um estado da arte em relação ao

reconhecimento de caracteres Kanji e a pertinência deste estudo na área de reconhecimento de padrões se encontra em WAKAHARA & KIMURA (1999). O artigo trata de problemas comuns em toda a área de reconhecimento de padrões tais como a semelhança entre características versus a semelhança entre padrões e conceitos como “boas” características, “boas” aproximações (*matching*), “boa” clusterização e aprendizado “real”. O trabalho dos autores neste estudo visa a obtenção de *insights* ao abordar um problema com este nível de complexidade e pretendem ter contribuído na área de análise de padrões e inteligência computacional.

No século 21, os avanços tecnológicos permitem a construção de sistemas de decisão mais sofisticados e robustos que em qualquer outra época. Atualmente a demanda por resultados satisfatórios vem frequentemente se sobrepondo a compreensão científica dos modelos. Em LEIGH & PURVIS e RAGUSA (2002) nota-se a preocupação dos autores em buscar soluções mais ecléticas que as tradicionais e usam mesmo o termo *românticas* ao definirem tais caminhos. A combinação de técnicas como o reconhecimento de padrões, redes neurais e algoritmos genéticos são abordagens consideradas pelos autores como alternativas.

Ao se verificar o estado da arte em aplicações de modelos semelhantes ao apresentado neste trabalho, deve-se buscar métodos que reúnem conceitos de conjuntos difusos com o de aprendizado automático. Esta síntese é bem representada nas redes *neuro-fuzzy*. Em RAMOS (1999) encontra-se uma definição para estas redes:

Uma rede neuro-fuzzy pode ser definida como sendo um sistema fuzzy que é treinado por algum algoritmo derivado da teoria das redes neurais.

A utilização de redes neuronais proporcionam a habilidade para reconhecer e classificar com robustez e capacidade para generalizar, por outro lado a utilização de técnicas difusas permitem uma boa compreensão do problema.

2.5 FAN – Free Associative Neurons

No ano de 1997, foi publicada a primeira abordagem acerca de FAN RAITTZ (1997-1), (1997-2), (1998). O centro da proposta do método é o desenvolvimento de uma estratégia de reconhecimento de padrões que garanta boa performance no aprendizado aliado às vantagens computacionais da clareza na representação dos padrões e portabilidade das unidades de representação que são chamadas neurônios ou FAN.

Na seqüência, apresentar-se-á a visão geral de FAN na época de sua origem, visando a subsidiar a compreensão do seu desenvolvimento e comparação com o seu estágio atual.

O restante deste tópico é fiel ao que é reportado em RAITTZ (1997-1).

“Em FAN, cada padrão de entrada é expandido em uma vizinhança difusa. Cada conjunto suporte dessa vizinhança é a combinação dos valores das características próximos dos originais.

O grau de similaridade entre a vizinhança difusa e o padrão original de entrada é feito por meio das técnicas utilizadas na teoria dos conjuntos difusos (KLIR, 1995).

O aprendizado acontece pela projeção da vizinhança difusa no espaço FAN. Há uma unidade FAN para cada classe do domínio do problema. Cada unidade é uma matriz composta por todas as combinações de características observadas em sua classe correspondente.

Durante o treinamento, cada combinação é representada por uma célula difusa que contém um peso correspondente à sua frequência de ocorrência e grau de pertinência. O treinamento é baseado no *reforço* nas células projetadas de FAN – se a classificação foi correta – ou em *esquecimento* (penalização) – se houve uma classificação incorreta.

(...)

Em FAN, dois princípios básicos fundamentam o processo de representação dos dados:

- um padrão representa mais que um simples ponto em termos de informação. Assim, na modelagem será considerada uma região ao redor do padrão;
- cada padrão contém mais informações que vão além dos valores individuais. Fazendo a correlação dos valores das características, a performance pode ser aumentada.

O primeiro aspecto está relacionado com a *decomposição do padrão*; ou seja, a geração de uma vizinhança em torno do padrão de entrada. A vizinhança constitui um conjunto de padrões difusos no espaço de entrada, próximos ao padrão original.

O outro passo no processo de modelagem dos dados é a *projeção de cada vizinhança difusa* em uma região denominada espaço FAN, que consiste em um conjunto de pontos do espaço \mathfrak{R}^{2H} que é comentado detalhadamente mais adiante. O algoritmo FAN mede a similaridade entre essas projeções e as unidades FAN.

(...)

Em suma, o processo de modelagem é realizado pelos seguintes procedimentos:

- decomposição do padrão (vizinhança difusa);
- projeção da vizinhança difusa.

Na decomposição, o padrão de entrada é transformado em um conjunto de padrões difusos. Dado um padrão de entrada \mathbf{x} , a seguinte expressão representa o conjunto de padrões difusos:

$$\bar{\mathbf{x}}_z = (\mu_{x_1}(v_1(l_1))/v_1(l_1), \dots, \mu_{x_p}(v_p(l_p))/v_p(l_p)) \quad (4.1)$$

onde:

$$z = 1, (2D + 1)^p;$$

D é o *raio* de decomposição. Ele estabelece o número de inteiros em cada lado da coordenada original x_i (*centro*) do vetor de entrada \mathbf{x} ;

$v_i(l_i)$ é o valor da i -ésima coordenada inteira, próxima ao valor original x_i . O índice $l_i = 1, 2D+1$, variando para a vizinhança de cada coordenada;

$\mu_{x_i}(v_i(l_i))$ é a função de pertinência que mede a similaridade entre o valor $v_i(l_i)$ e o original x_i ($i=1, p$).

(...)

Uma vez criados os vizinhos difusos, o processo de modelagem dos dados dá início à projeção da vizinhança. A idéia é projetar o padrão original \mathbf{x} sobre o espaço de combinações das características.

Sua dimensão é igual ao produto cartesiano difuso da vizinhança das características, em que a similaridade em relação aos valores originais é obtida por meio de técnicas difusas. Isso é feito projetando cada vizinho difuso $\bar{\mathbf{x}}_z$ sobre o espaço cartesiano difuso, gerando a projeção \mathbf{F}_z .

O primeiro parâmetro a ser ajustado é o *nível de combinação das características*, denominado H . Este fator tem impacto direto na dimensão do espaço FAN. Quanto maior for H , maior se torna a dimensão do espaço FAN.

O escopo de valores de H é o intervalo $[1, p]$, onde p é a dimensão do espaço de entrada.

Dado um vetor difuso de vizinhança como em (1), há p características a serem combinadas em H -uplas. O conjunto destas combinações formam a projeção.

Desta forma, um fator $H = 2$ implica que as características serão combinadas em pares. A projeção completa \mathbf{F}_z é formada por todas as $C_p^H + p$ combinações de características. Esta combinação é composta pelos índices dos vetores $\bar{\mathbf{x}}_z$ e pelos valores das coordenadas $v_j(l_j)$ a serem combinados.

Cada combinação é avaliada pelo produto dos graus de pertinência correspondentes $\mu_{x_i}(v_i(l_i))$. A projeção de \mathbf{F}_z é determinada então por:

$$\mathbf{F}_z \in \mathfrak{r}^{2H(C_p^H + p)} = \text{Proj}(\bar{\mathbf{x}}_z) = (f_m^z, \mathbf{a}, \boldsymbol{\alpha}) \quad (4.2)$$

onde $\mathbf{a} = (a_1, \dots, a_H)$ é o vetor dos índices, com $1 \leq a_j \leq p$, $\boldsymbol{\alpha} = (V_1(i_1), \dots, V_H(i_H))$ e

$$f_m^z = \prod_{j=1, H} \mu_{x_i}(V_i(a_j)), \text{ com } m = 1, (C_p^H + p). \quad (4.3)$$

(...)

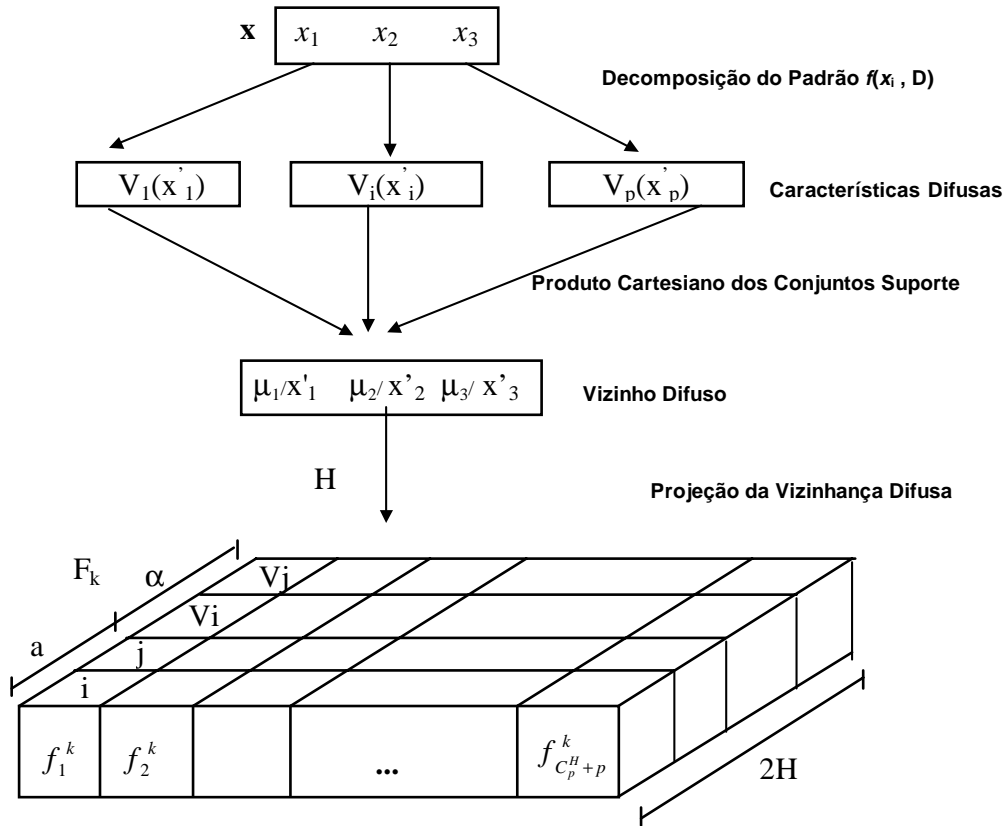
Método de Aprendizado FAN

As expressões (4.2) e (4.3) descrevem como um vizinho difuso é representado em um subespaço do FAN. Na implementação do aprendizado, entretanto, deve-se considerar não apenas um vizinho. Na expressão abaixo, \mathbf{F} é calculado usando-se todos os vizinhos possíveis para H e D definidos para a aplicação.

Em alguns casos práticos, não é usada a totalidade, mas sim uma quantidade que seja representativa, evitando problemas de explosão combinatorial. O número dessas combinações é usado como parâmetro nas aplicações e recebe o nome de *número de combinações das características*. A *projeção da vizinhança difusa* é a união dos vizinhos difusos projetados de \mathbf{F}_z :

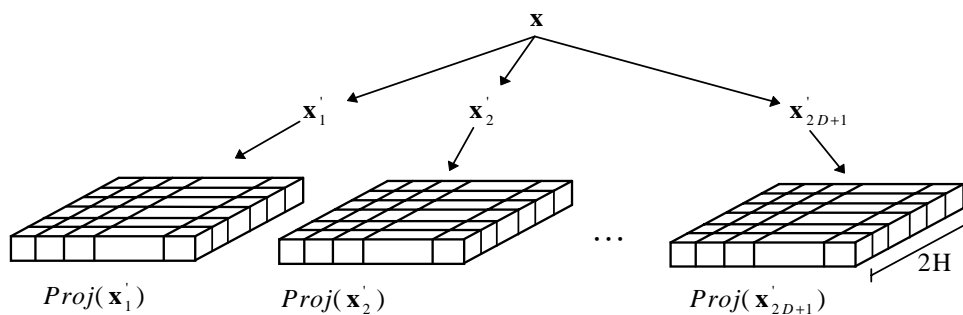
$$\mathbf{F} = \bigcup_{z=1, (2D+1)^p} \mathbf{F}_z. \quad (4.4)$$

Figura 4.1 – Visão esquemática do processo de modelagem de dados



A projeção completa da vizinhança difusa \mathbf{F} é um conjunto de $(2D+1)^p$ cubos similares àqueles, na parte inferior da *Figura 4*. A *Figura 5* representa a união mostrando cada elemento \mathbf{F}_z do conjunto resultante \mathbf{F} .

Figura 4.2 – Projeção da vizinhança difusa



O aprendizado envolve a comparação entre a projeção \mathbf{F} e todas as unidades \mathbf{FAN}^j ($j=1, C$, onde C é o número de classes do domínio do problema).

A dimensão de \mathbf{F} é definida pela expansão do padrão de entrada \mathbf{x} (i.e., depende dos valores das coordenadas x_i de \mathbf{x} e dos parâmetros D, H). \mathbf{FAN}^j ,

entretanto, possui uma dimensão maior, incluindo todas as combinações de características que ocorrem para todo $\mathbf{x}^k \in \mathbf{FAN}^j$, cuja comparação é feita avaliando o grau ao qual \mathbf{F} é um subconjunto de \mathbf{FAN}^j .

O cálculo da similaridade entre \mathbf{F} e \mathbf{FAN}^j bem como as expressões usadas para reforço e esquecimento não representam a essência do método proposto, podendo variar entre uma aplicação e outra.

A similaridade entre \mathbf{F} e \mathbf{FAN}^j é dada por:

$$S(\mathbf{F}, \mathbf{FAN}^j) = 1 - \prod_g \left(1 - \frac{f_g \cdot fan_g^j}{fan^j} \right) \quad (4.5)$$

onde fan_g^j é uma célula do conjunto difuso \mathbf{FAN}^j ; f_g é uma célula do conjunto difuso \mathbf{F} (obtido pela Expressão 4.3); g é o índice que cobre todo o domínio de \mathbf{FAN}^j que possui um correspondente em \mathbf{F} e fan^j é o número total de ocorrências na unidade \mathbf{FAN}^j .

A Expressão (5) mede a similaridade da projeção \mathbf{F} em cada unidade \mathbf{FAN} , \mathbf{FAN}^j .

A maior similaridade indica a classe associada ao padrão de entrada \mathbf{x} . Em outras palavras, a saída da rede é j^* -ésima classe, onde $j^* \in [1, C]$ é o índice da unidade \mathbf{FAN} que apresenta maior similaridade, \mathbf{FAN}^j . Isto é obtido por:

$$S(\mathbf{F}, \mathbf{FAN}^{j*}) = \max_j S(\mathbf{F}, \mathbf{FAN}^j) \quad (4.6)$$

Uma vez que a saída da rede é determinada, acontece o aprendizado (*backward computation*).

Primeiramente ocorre a checagem do acerto na classificação ($j^* = j^f$, onde j^f é a classe do domínio associada ao padrão de entrada \mathbf{x}) ou erro de classificação, em caso contrário.

Em ambos os casos, a mudança dos valores de pertinência de \mathbf{FAN}^{j*} é determinada pelo grau de pertinência da célula f_g .

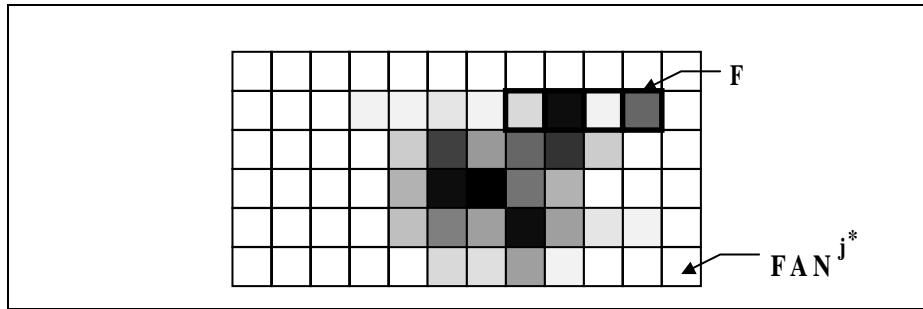
Quando a resposta de \mathbf{FAN} for correta, o procedimento consiste no fortalecimento das células \mathbf{FAN}^{j*} interceptadas pela projeção \mathbf{F} , que consiste em somar ao conteúdo atual da célula fan_g^j o valor de pertinência f_g .

Por sua vez, a pertinência será determinada pela somatória das pertinências na célula dividida pelo total de ocorrências na unidade: $\frac{fan_g^j}{fan^j}$.

No caso de ocorrer uma classificação errada, o algoritmo realiza um procedimento de *esquecimento*, diminuindo o valor das células FAN^j interceptadas pela projeção **F**. Nesse caso, a pertinência f_g é subtraída do conteúdo da célula. Se este conteúdo tornar-se negativo após a subtração, seu valor será definido como zero.

A *Figura 4.3* representa a idéia que se encontra por trás dos procedimentos de medidas de similaridade, *reforço* e *esquecimento*. Quanto mais escuras forem as células, maior é o grau de pertinência em relação ao conjunto difuso (**F** ou FAN^j).

Figura 4.3 – Intersecção Entre **F** e FAN^j



4.5 Algoritmo FAN

Conjunto de Treinamento: $P_k(\mathbf{x}^k, \mathbf{t}^k)$, ($k = 1, N$), tal que $\mathbf{x}^k \in \mathbf{r}^p$ e $\mathbf{t}^k \in \mathbf{r}^q$

onde: \mathbf{x}^k é o k-ésimo padrão a ser aprendido; \mathbf{t}^k indica a classe à qual \mathbf{x}^k pertence; \mathbf{r}^p é o espaço das características dos padrões; \mathbf{r}^q é o espaço dos padrões e N é o número de dados de treinamento.

Inicialização

Faça $FAN^j = 0$ para $j = 1, C$, onde C é o número de classes no domínio do problema.

Ajuste o nível de combinações de características H e o raio de decomposição D , tal que $1 \leq H \leq p$, e $D > 0$.

Para cada par $P_k(\mathbf{x}^k, \mathbf{t}^k)$, faça o seguinte:

Primeiro passo: decomposição do padrão

Para cada coordenada x_i^k do vetor \mathbf{x}^k , gere as seqüências $V_i^k = [\text{int}(x_i^k - D), \dots, \text{int}(x_i^k), \dots, \text{int}(x_i^k + D)]$, para $i=1, p$.

Escolha uma função de pertinência $\mu_{x_i^k}(x)$ tal que $\mu_{x_i^k}(x < x_i^k - D) = 0$, $\mu_{x_i^k}(x_i^k) = 1$, $\mu_{x_i^k}(x > x_i^k + D) = 0$.

Crie os vizinhos difuso $\bar{\mathbf{x}}_z^k \in \mathbf{r}^p$, onde $z = 1, (2D+1)^p$ assumindo valores nas seqüências V_i^k como conjunto suporte, *i.e.*, $\bar{\mathbf{x}}_z^k = \mu_{x_1^k}(V_1^k(l_1))/V_1^k(l_1), \dots, \mu_{x_p^k}(V_p^k(l_p))/V_p^k(l_p)$, onde cada $V_i^k(l_i)$ é um membro da seqüência V_i^k , o qual $1 \leq l_i \leq (2D+1)$, $i=1, p$.

Segundo passo: projeção da vizinhança difusa

(i) *projeção do vizinho*: Para cada número difuso $\bar{\mathbf{x}}_z^k$, $z = 1, (2D+1)^p$, encontre a projeção $\mathbf{F}_z \in \mathbf{r}^{2H(C_p^H + p)} = \text{Proj}(\bar{\mathbf{x}}_z^k) = (f_m^z, \mathbf{a}, \boldsymbol{\alpha})$; \mathbf{a} é o vetor de índices $\mathbf{a} = (a_1, \dots, a_H)$, sendo $1 \leq a_j \leq p$, $\boldsymbol{\alpha} = (V_1^k(i_1), \dots, V_H^k(i_H))$, $f_m^z = \prod_{j=1, H} \mu_{x_i^k}(V_i^k(a_j))$, com $m = 1, (C_p^H + p)$.

(ii) *projeção da vizinhança*: calcule a projeção da vizinhança difusa \mathbf{F}^k tal que $\mathbf{F}^k = \bigcup_{z=1, (2D+1)^p} \mathbf{F}_z$.

Terceiro passo: grau de similaridade

$S(\mathbf{F}, \mathbf{FAN}^j) = 1 - \prod_g \left(1 - \frac{f_g \cdot fan_g^j}{fan_g^j} \right)$, onde g é o índice que cobre todo o domínio de

FAN, correspondente ao domínio de \mathbf{F}^k e fan_g^j é o número total de ocorrências de reforços na unidade FAN^j.

Quarto passo: saída da Rede

Encontre FAN^{j*}, tal que $S(\mathbf{F}^k, \mathbf{FAN}^{j*}) = \max_j S(\mathbf{F}^k, \mathbf{FAN}^j)$.

Faça j^* ser a saída da rede.

Quinto passo: retroprocessamento

Escolha o fator $\lambda = f_g$, como parâmetro de reforço e aplique a expressão [7] para $FAN^{j^{tk}}$.

Verifique se $j^* = j^{tk}$, onde j^{tk} é a classe domínio associada ao objetivo t^k .

Escolha o fator λ de acordo com a classificação da rede: se a classificação for correta, ou seja, $j^* = j^{tk}$, escolha λ como parâmetro de *reforço*.

Se houve uma classificação incorreta, tome $\lambda = -f_g$, como o parâmetro de *esquecimento*.

Em ambos os casos, utiliza a seguinte expressão FAN^{j^*} :

$$fan_g^{NEW z^*} = fan_g^{OLD z^*} + \lambda \quad (4.7)$$

4.6 Conclusão

FAN é baseado no aprendizado conexionista e modelagem difusa. Os principais passos para a realização do reconhecimento consistem em: (i) gerar vizinhança do padrão; (ii) gerar as combinações das características que compõem os vizinhos; e (iii) testar a semelhança entre a projeção assim obtida e cada classe do domínio do problema, representada por unidades FAN alterando-as conforme os resultados.

Ainda que tenha sido mostrado um algoritmo com expressões bem definidas e para o espaço completo das projeções, os pressupostos acima podem ser mantidos, alteradas as funções de similaridade e armazenamento, bem como o domínio das projeções pode ser reduzido, tomando-se, em vez de todos os vizinhos e combinações de H, apenas uma parte representativa da mesma. Isto foi feito na maior parte das aplicações práticas.”

No modelo proposto nos próximos capítulos mostra-se uma abordagem *neuro-fuzzy* para classificação de padrões. A geração das funções de pertinência dos padrões em cada classe é realizada automaticamente de forma semelhante a utilizada nas redes neuronais. A manutenção dos conjuntos

difusos em cada neurônio permitem a visualização dos resultados obtidos para a compreensão do problema tratado.

3 ESTRUTURA ATUAL DO MODELO FAN COMO MÉTODO PARA RECONHECIMENTO DE PADRÕES

O foco principal da atenção na elaboração do modelo proposto busca encontrar uma forma de conseguir um método de RP dedicado à implementação computacional, que enfatize o aspecto da estrutura das informações envolvidas, mais que algoritmos matemáticos complexos.

Foram realizadas várias modificações do modelo desde a sua publicação em RAITTZ (1997-1).

Neste capítulo, é apresentada a estrutura do modelo da forma que ele foi construído nas suas últimas aplicações com as simplificações e adaptações que ele sofreu desde então.

No que tange às simplificações, pode-se ressaltar a eliminação do parâmetro de combinações de características, chamado H , em RAITTZ (1997).

Ao se considerar tal parâmetro, tinha-se então por objetivo aumentar o número de informações a serem relacionadas aos padrões. No entanto, observa-se na prática que a qualidade no reconhecimento devido à adoção do parâmetro $H > 1$ não é aumentada, enquanto o tempo computacional demandado nesse caso o é de forma significativa. Dessa forma, optou-se por descrever o modelo sem levá-lo em conta, tornando-o mais simples e compreensível.

Já no aspecto das adaptações do método, pode-se listar:

- a) a alteração na forma de realizar a penalização (esquecimento). No modelo antigo, a penalização se fazia por meio da subtração de um valor nos intervalos difusos vizinhos às características em um FAN quando ocorresse

um erro de reconhecimento. No modelo atual, considerou-se que o aprendizado ocorre de forma mais adequada caso realize um *desgaste* nas mesmas regiões através do produto por um fator próximo de 1 ($1-\alpha$, $\alpha \rightarrow 0$), a ser considerado de forma difusa, tanto mais próximo de um quanto mais distante do valor real da característica considerada no padrão classificado equivocadamente. Além de ser mais eficaz, esta forma de tratamento é mais razoável se considerar-se que ela pode ser tomada como uma interseção na teoria dos conjuntos difusos;

- b) a inicialização dos neurônios com valores aleatórios. No modelo anterior, as unidades FAN eram inicializadas com valores nulos para todas as células. Ao se utilizarem valores aleatórios iniciais para as células, pode-se melhorar a capacidade de adaptação dos neurônios em função da penalização, que terá lugar desde a avaliação do primeiro padrão, o que pode não ocorrer no outro caso;
- c) a adequação do modelo aos conceitos da teoria dos conjuntos difusos. A descrição do modelo atual visa a adequar os princípios de FAN aos daquela teoria, o que pode ser verificado nesse capítulo;
- d) a utilização de um *hedge* no teste de um padrão. A experiência indicou como funcional a utilização de um *hedge* de dilatação na t-conorma que define o grau de pertinência de um padrão em relação a uma determinada unidade FAN, sendo utilizado para tal a raiz quadrada conforme o passo 5 do algoritmo.

Além disso, uma implementação alternativa do método foi feita por DANDOLINI (1999), sobre a qual não se aprofunda neste trabalho e é apenas comentada no capítulo de aplicações.

Várias foram as técnicas estudadas que subsidiaram a construção do método tal como ele se encontra neste estágio.

A discussão de cada uma delas está fora do escopo deste trabalho.

Na seqüência, é apresentado o modelo tal como ele se encontra por meio de seu algoritmo.

3.1 Problema

Classificar um indivíduo “I” em uma de i classes, considerando-se C características, sendo dado um conjunto de H (não confundir com o parâmetro de combinação de características de FAN 1997) observações para treinamento.

3.2 Notação

- X_{hc}^{tr} ; $c = 1, C$; $h = 1, H$; é o valor da característica “c” do h-ésimo indivíduo do conjunto de treinamento (tr).

- X_{lc}^{te} ; $c = 1, C$; $l = 1, L$; é o valor da característica “c” do l-ésimo indivíduo do conjunto de teste (te).

Os valores de um conjunto de X_{hc}^{tr} formam um padrão:

$$X_{hc}^{tr} = [X_{h1}^{tr}, X_{h2}^{tr}, \dots, X_{hc}^{tr}] \quad c = 1, C$$

3.3 Estrutura dos dados de treinamento

Tabela 1 – Estrutura dos dados de treinamento

Indivíduo	Características				Classe
	1	2	...	C	$p_i ; i = 1, l$
1	X_{11}^{tr}	X_{12}^{tr}	...	X_{1C}^{tr}	$p_i ; i = 1, l$
2	X_{21}^{tr}	X_{22}^{tr}	...	X_{2C}^{tr}	$p_i ; i = 1, l$
...
H	X_{H1}^{tr}	X_{H2}^{tr}	...	X_{HC}^{tr}	$p_i ; i = 1, l$

3.4 Estrutura dos dados de teste

Tabela 2 – Estrutura dos dados de teste

Indivíduo	Características				Classe
	1	2	...	C	$p_i ; i = 1, l$
1	X_{11}^{te}	X_{12}^{te}	...	X_{1C}^{te}	$p_i ; i = 1, l$
2	X_{21}^{te}	X_{22}^{te}	...	X_{2C}^{te}	$p_i ; i = 1, l$
...
L	X_{L1}^{te}	X_{L2}^{te}	...	X_{LC}^{te}	$p_i ; i = 1, l$

p_i indica a classe “i”, e “p” significa que essa classe foi determinada para um dado padrão.

3.5 Processo

O processo para classificação consiste em duas etapas; a primeira é de treinamento e a segunda de teste.

3.6 Treinamento

A etapa de treinamento pode ser compreendida nos seguintes passos:

3.6.1 Passo 1: Normalização

Para cada característica “c”, a normalização é realizada pela definição de um novo valor para X_{hc}^{tr} , pela substituição do valor original por \hat{X}_{hc}^{tr} , definido como:

$$\hat{X}_{hc}^{tr} = \frac{X_{hc}^{tr}}{\max_h X_{hc}^{tr}} \quad c = 1, C \quad h = 1, H$$

Observe-se que o valor máximo que \hat{X}_{hc}^{tr} assumirá será igual a 1.

A normalização nesse contexto não diz respeito à normalização de conjuntos difusos, que significa tornar a altura do conjunto difuso igual à unidade.

3.6.2 Passo 2: Determinação do FAN inicializador

O neurônio FAN inicializador para cada uma das classes i , $i = 1, I$ (FAN_i), denominado de Y_{icj} , será determinado por meio da geração de números aleatórios no intervalo $[0, 1]$.

Assim, para cada classe “i”, gera-se matrizes de dimensão $C \times J$ de números aleatórios:

$Y_{icj} \quad i = 1, I \text{ classes}$

$c = 1, C \text{ características}$

$j = 1, J \text{ números definidos para o suporte dos conjuntos difusos}$

3.6.3 Passo 3: Construção dos conjuntos difusos para os padrões

Constrói-se para o primeiro padrão (indivíduo), $c = 1$, C conjuntos difusos, que representam números difusos dos valores particulares observados no padrão, e já normalizados:

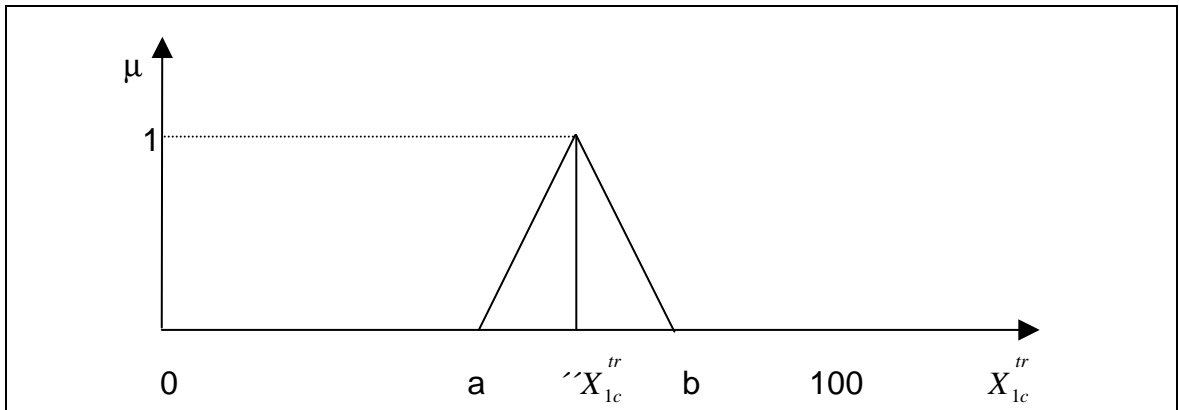
$$[{}^trX_{11}, {}^trX_{12} \dots {}^trX_{1c}] \quad c = 1, C$$

Como o universo de discurso está em $[0, 1]$ torna-se

$${}^trX_{1c} = {}^trX_{1c} \times J$$

Assim, ${}^trX_{1c}$ será o centro do número difuso, cujo universo de discurso é definido no intervalo $[0, J]$, e cujo suporte é $[a, b]$, conforme a *Figura 4*:

Figura 4 – Representação de um número difuso em torno de uma característica



Onde:

$$a = \text{int}({}^trX_{1c}) - d + 1, \text{ e}$$

$$b = \text{int}({}^trX_{1c}) + d$$

$\text{int}(x)$ é o valor da parte inteira do número, e $d-1$ é igual ao número de vizinhos inteiros de cada lado do valor central, cujos graus de pertinência são maiores que zero.

Os conjuntos difusos são definidos como:

$$\mu(X_{1c}^{tr}) = \begin{cases} 0 & \text{se } X_{1c} \leq a \text{ ou } X_{1c} \geq b \\ \frac{X_{1c} - a}{\hat{X}_{1c}^{tr} - a} & \text{se } a \leq X_{1c} \leq \hat{X}_{1c}^{tr} \\ \frac{b - X_{1c}}{b - \hat{X}_{1c}^{tr}} & \text{se } \hat{X}_{1c}^{tr} \leq X_{1c} \leq b \end{cases}$$

No intervalo $[a, b]$, haverá 2 (d-1) números inteiros com graus de pertinência maiores que zero.

Esses graus de pertinência serão denotados por $\mu_{icj}|_{h=1}$.

Observe-se que j varia de zero a J , mas μ_{icj} só assumirá valores diferentes de zero no intervalo $[a, b]$, mesmo em a e b o valor de μ_{icj} será igual a zero.

3.6.4 Passo 4: Normalização dos graus de pertinência

Também aqui normalização não diz respeito a tornar a altura do conjunto difuso igual à unidade. Os graus de pertinência μ_{icj} serão normalizados definindo-se $\hat{\mu}_{icj}$ por:

$$\hat{\mu}_{icj}|_{h=1} = \frac{\mu_{icj}|_{h=1}}{\sum_j \mu_{icj}|_{h=1}} \quad \text{para } c = 1, C \text{ e } j = 1, J$$

observe-se que $\sum_j \mu_{icj}|_{h=1}$ é a cardinalidade do conjunto \hat{X}_{1c}^{tr}

3.6.5 Passo 5: Determinação da média geométrica (*hedge*) do FAN_i e dos graus de pertinência modificados do padrão

Para cada valor inteiro no intervalo [a, b], pode-se calcular a média geométrica entre \hat{Y}_{icj} e $\mu_{icj|_{h=1}}$, onde

$$\hat{Y}_{icj} = \frac{Y_{icj}}{\sum_j Y_{icj} + \Delta_{ic}}$$

\hat{Y}_{icj} são os valores normalizados do neurônio FAN, e Δ_{ic} é o total de penalizações para a célula.

Fora desse intervalo, os valores serão iguais a zero.

Essas médias geométricas determinam novos valores de graus de pertinência que serão denotados por $\hat{\mu}_{icj|_{h=1}}$.

Observe-se que aqui $j = 1, J$ significa que a normalização está sendo feita a partir de todos os valores inteiros no universo de discurso, uma vez que \hat{Y}_{icj} assumirá valores em todo o universo, e não somente no intervalo [a, b], de tal forma que a cardinalidade do conjunto $\hat{\hat{Y}}_{icj}$ é no máximo igual a 1.

Ainda assim, para o cálculo da média geométrica, só importarão os valores de $\hat{\hat{Y}}_{icj}$ no intervalo de [a, b], dado que somente aí $\hat{\mu}_{icj|_{h=1}} \neq 0$.

As médias geométricas $\hat{\mu}_{icj|_{h=1}}$ representam a t-norma entre os dois graus de pertinência, e são dados por:

$$\hat{\mu}_{icj|_{h=1}}^i = (\hat{\mu}_{icj|_{h=1}} \times \hat{Y}_{icj})^{\frac{1}{2}} \quad i = 1, I \quad c = 1, C \quad j = 1, J$$

Para cada “i”, cada “c” e cada “j”, obter-se-á I x C x J t-normas, mas diferentes de zero somente no intervalo [a, b].

3.6.6 Passo 6: Determinação da força de representação do padrão

Tomando-se a característica “c” da classe “i”, determinou-se J t-normas $\mu_{icj|_{h=1}}^i$. Pode-se calcular o produto das raízes quadradas dos complementos dessas t-normas, definindo-se:

$$\prod_j (1 - \mu_{icj|_{h=1}}^i)^{\frac{1}{2}}$$

O complemento desse produto é a t-conorma. Sugerindo uma nova métrica, podemos dizer que:

$$\mu_{ic|_{h=1}}^u = 1 - \prod_j (1 - \mu_{icj|_{h=1}}^i)^{\frac{1}{2}} \quad i = 1, I \quad c = 1, C$$

Tomando-se agora a classe “i”, pode-se determinar:

$$\prod_c (1 - \mu_{ij|_{h=1}}^u)^{\frac{1}{2}}$$

E o seu complemento:

$$\mu_{i|_{h=1}}^u = 1 - \prod_c (1 - \mu_{ic|_{h=1}}^u)^{\frac{1}{2}} \quad i = 1, I$$

Verifica-se então o maior valor de $\mu_{i|_{h=1}}^u$ para cada i, i = 1, I.

Esse valor informa qual a classe esperada do padrão X_{1c}^{tr} .

A métrica apresentada neste trabalho para os operadores difusos não é aqui justificada. Sua prova pode resultar em novas discussões e trabalhos futuros.

3.6.7 Passo 7: Tomada de decisão

O padrão X_{lc}^{tr} pertence a classe pi , $i=1, l$.

Depois de treinado, e segundo o critério acima descrito, ele será classificado na classe $^{p'}i$, $i = 1, l$.

Se $i = ^{p'}i$ reforça-se a classe $^{p'}i$; caso contrário reforça-se a classe pi e penaliza-se a classe $^{p'}i$.

3.6.8 Passo 8: Penalização e/ou reforço

Reforço: O reforço é realizado na classe pi .

Toma-se para o indivíduo X_{lc}^{tr} ,

$$\gamma_{icj}^R = \gamma_{icj}^{\mu} + \gamma_{icj}, \text{ para } i = ^pi, j = 1, J, c = 1, C$$

Penalização: A penalização é realizada na classe $^{p'}i$:

Toma-se para o indivíduo X_{lc}^{tr}

$$\gamma_{icj}^P = 1 - \alpha * (\gamma_{icj}^{\mu}), \quad i = ^{p'}i, \quad j = 1, J, c = 1, C$$

$$\gamma_{icj}^P = \gamma_{icj}^{\mu P} * \gamma_{icj}$$

Dessa forma o novo valor da fan penalizado é:

$$\gamma_{icj}^P = (1 - \alpha * (\gamma_{icj}^{\mu})) * \gamma_{icj}, \quad i = ^{p'}i, \quad j = 1, J, c = 1, C$$

O treinamento é realizado repetindo-se o procedimento realizado a partir do passo 3 para todo $h = 1, H$. Todo o procedimento pode ser realizado mais de uma vez, ou seja, em varias épocas, conforme estratégia a ser definida.

3.6.9 Passo 9: Teste

O teste consiste em verificar para um novo conjunto de padrões, conjunto de teste $l=1, L$, a capacidade do método de indicar a classe correta para cada padrão, utilizando-se as expressões (1) e (2).

Essa capacidade é expressa em termos percentuais por meio da taxa de erro, “ t_e ”, dada por:

$$t_e = \frac{L - \sum U(l)}{L} * 100\%$$

$$\text{onde } U(l) = \begin{cases} 1 & \text{se } p_i = p'_i \\ 0 & \text{se } p_i \neq p'_i \end{cases}$$

Desse modo, quanto maior a capacidade de reconhecimento, menor será a taxa de erro.

Os gráficos apresentados a seguir representam os resultados dos treinamentos em uma aplicação real onde são consideradas sete classes i. e. sete neurônios para um problema de classificação de padrões com três características.

Cada gráfico se refere a um neurônio FAN que representa uma característica do domínio do problema.

Gráfico 1 – Representação do neurônio correspondente à primeira classe em uma aplicação prática após o treinamento

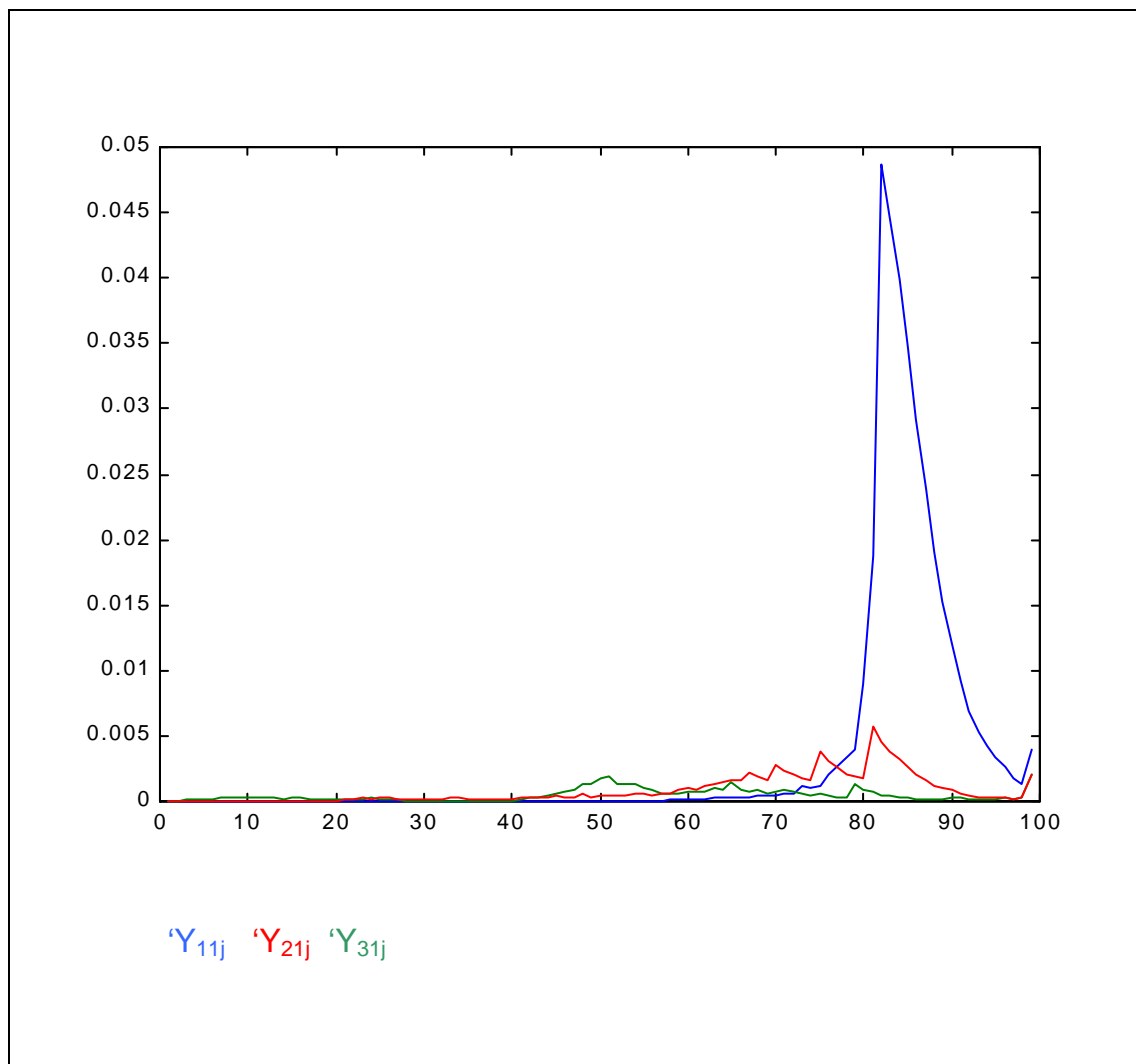


Gráfico 2 – Representação do neurônio correspondente à segunda classe em uma aplicação prática após o treinamento

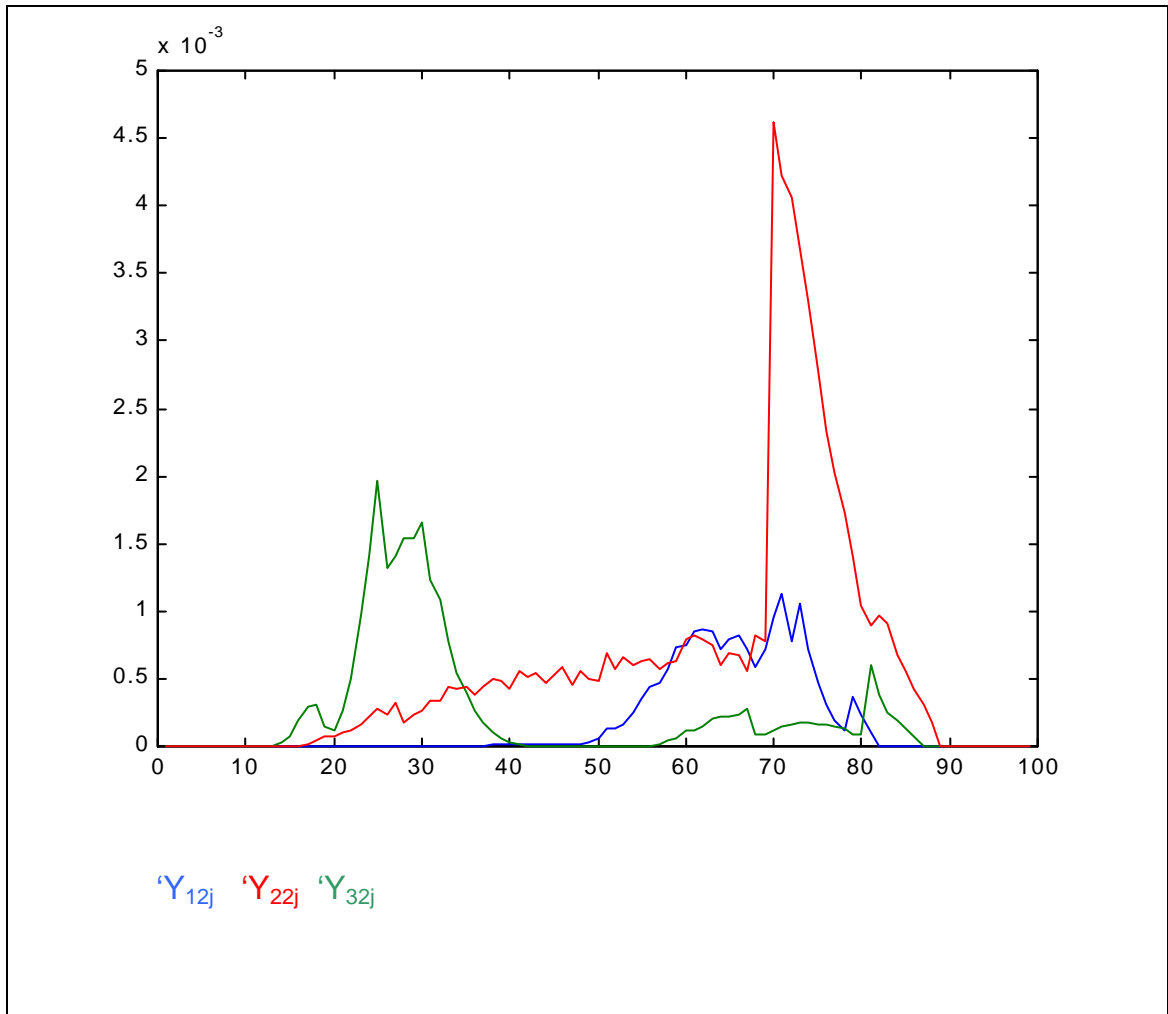


Gráfico 3 – Representação do neurônio correspondente à terceira classe em
uma aplicação prática após o treinamento

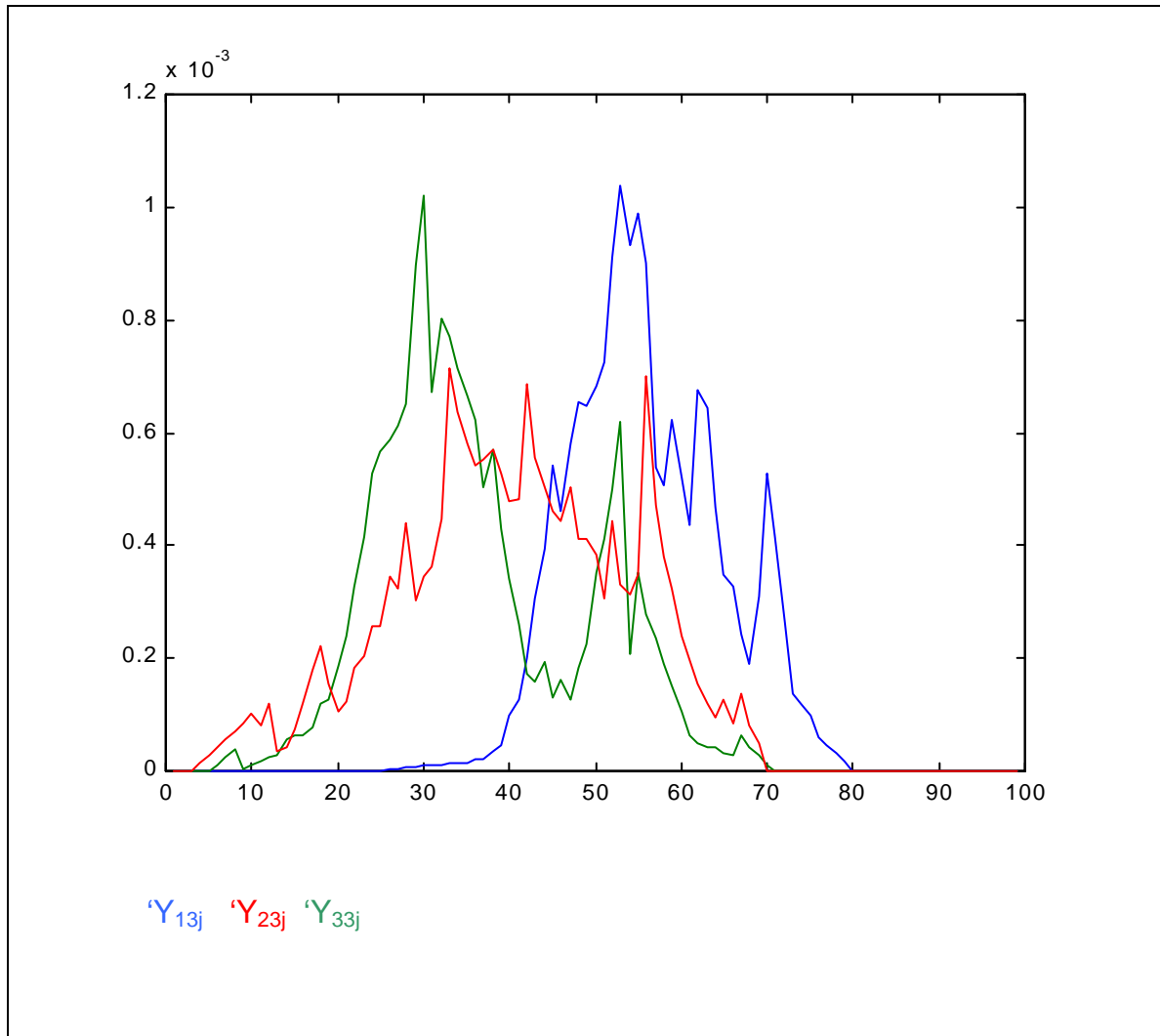


Gráfico 4 – Representação do neurônio correspondente à quarta classe em uma aplicação prática após o treinamento

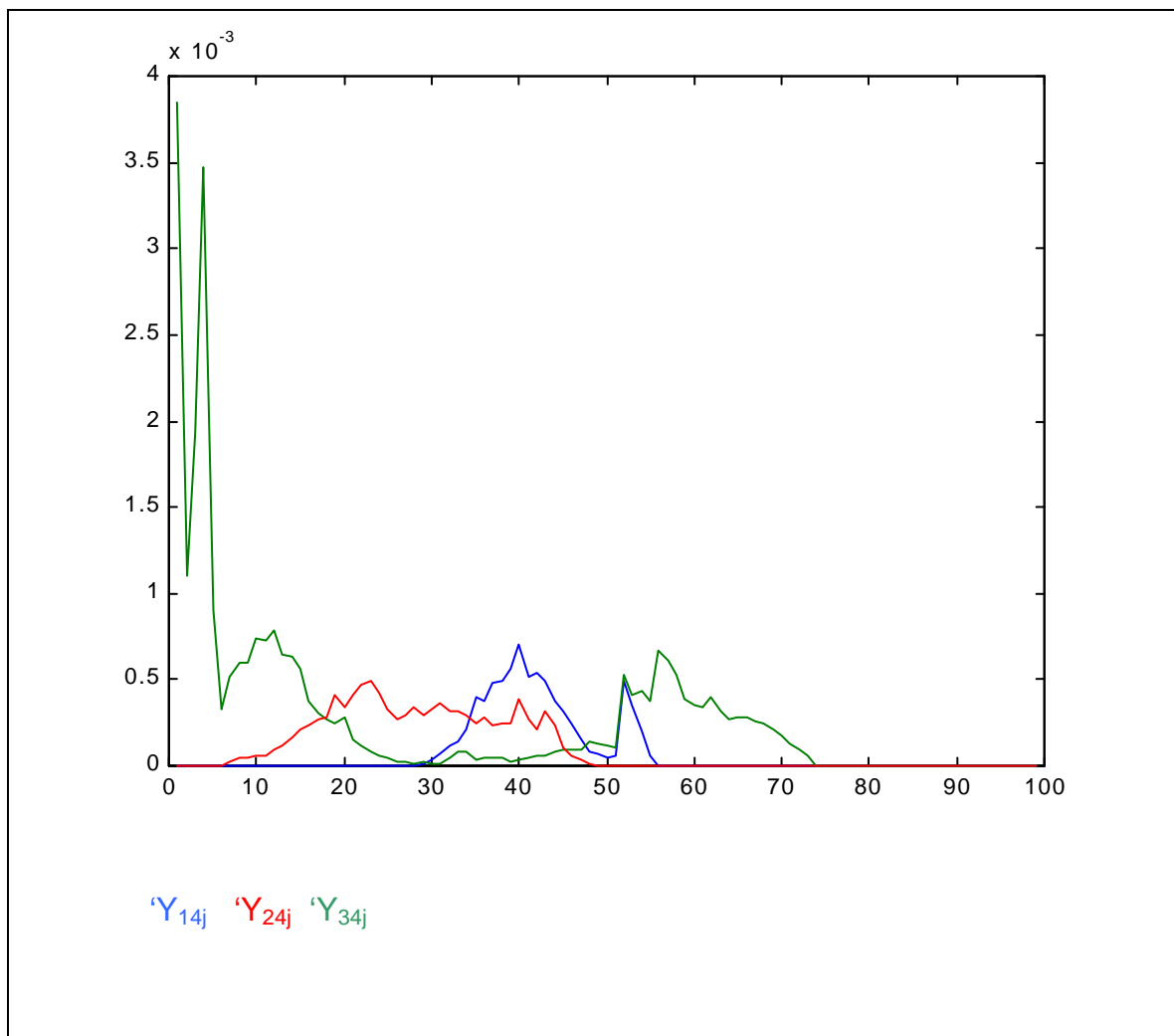


Gráfico 5 – Representação do neurônio correspondente à quinta classe em
uma aplicação prática após o treinamento

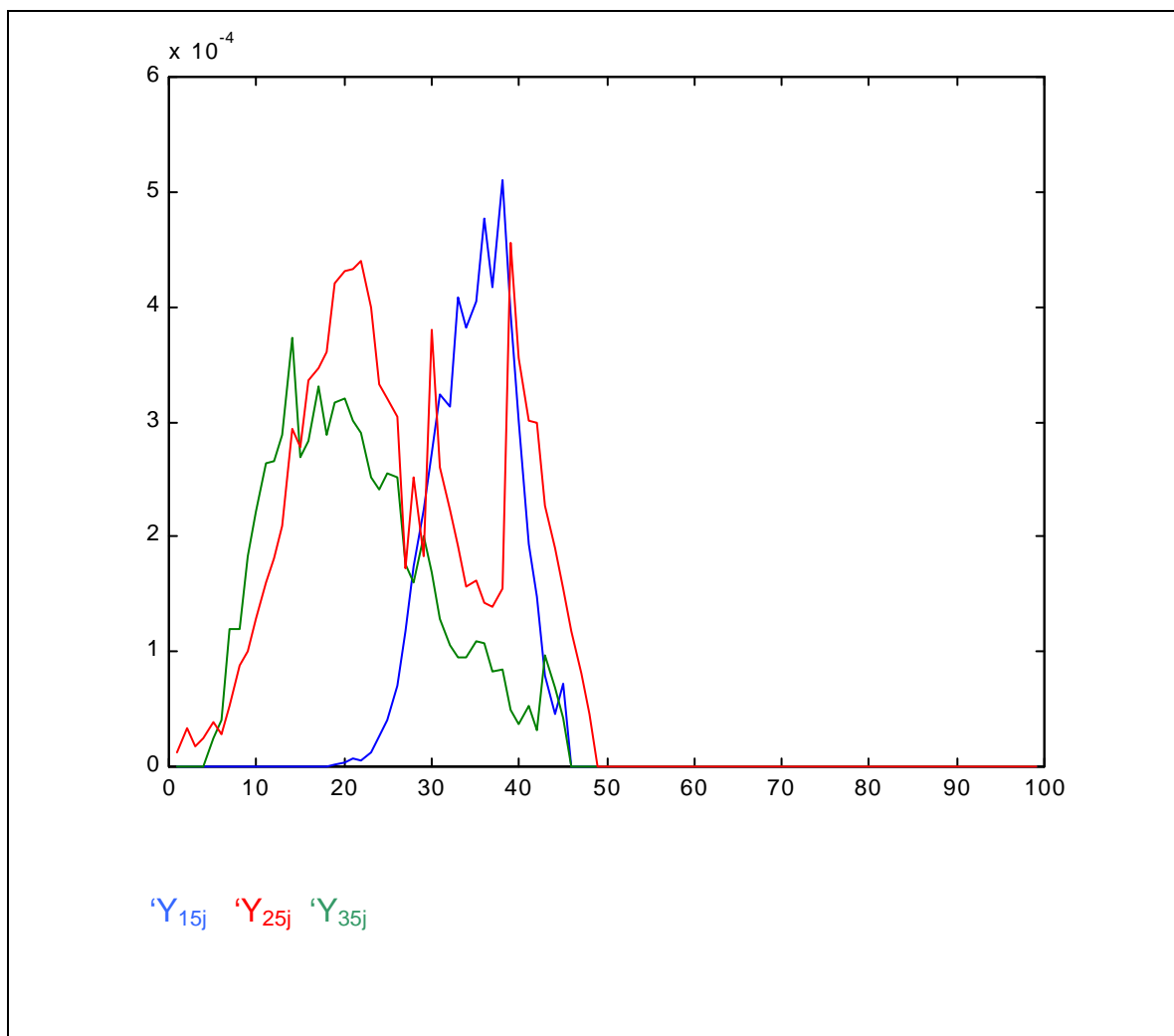


Gráfico 6 – Representação do neurônio correspondente à sexta classe em uma aplicação prática após o treinamento

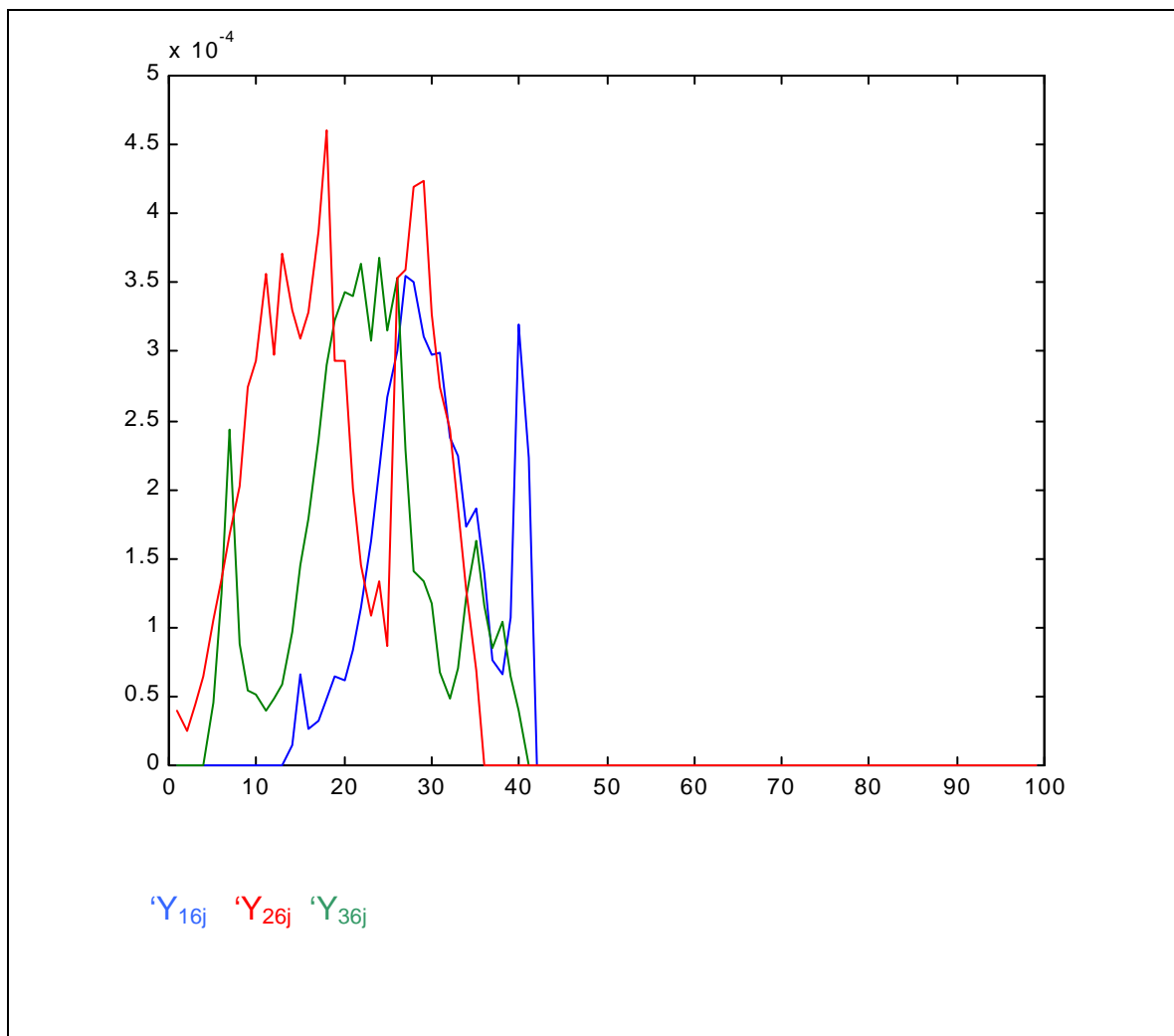
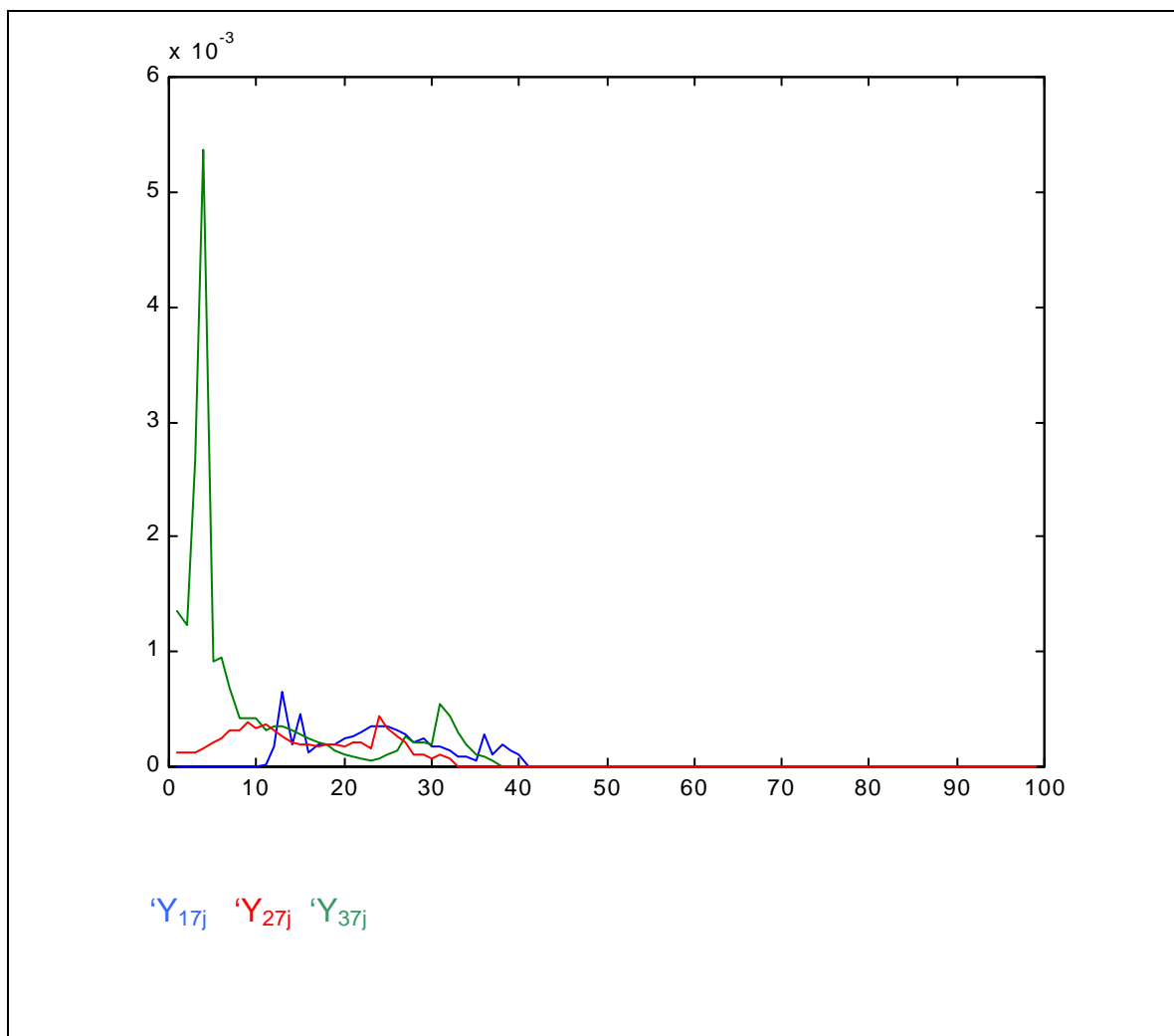
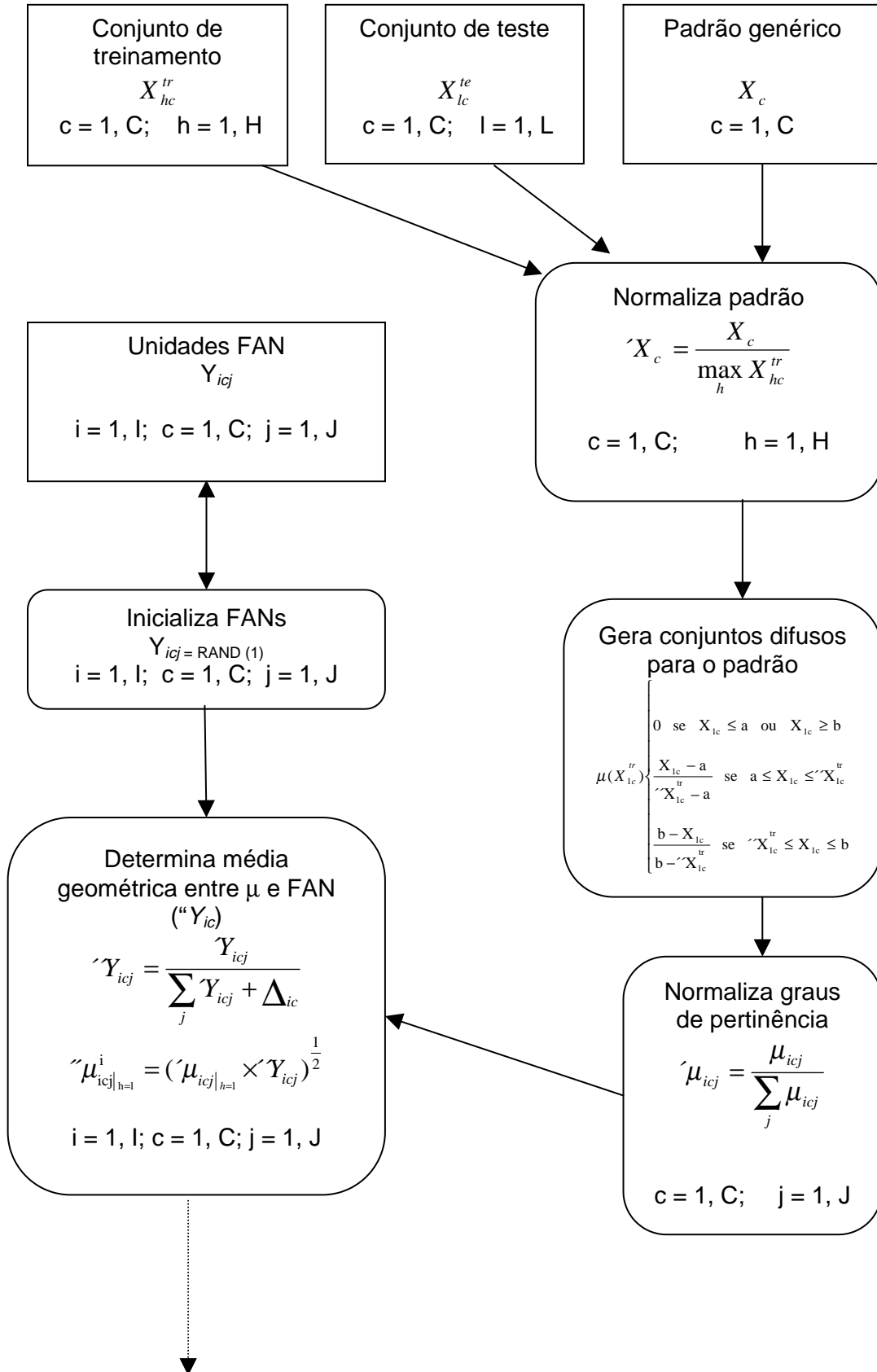
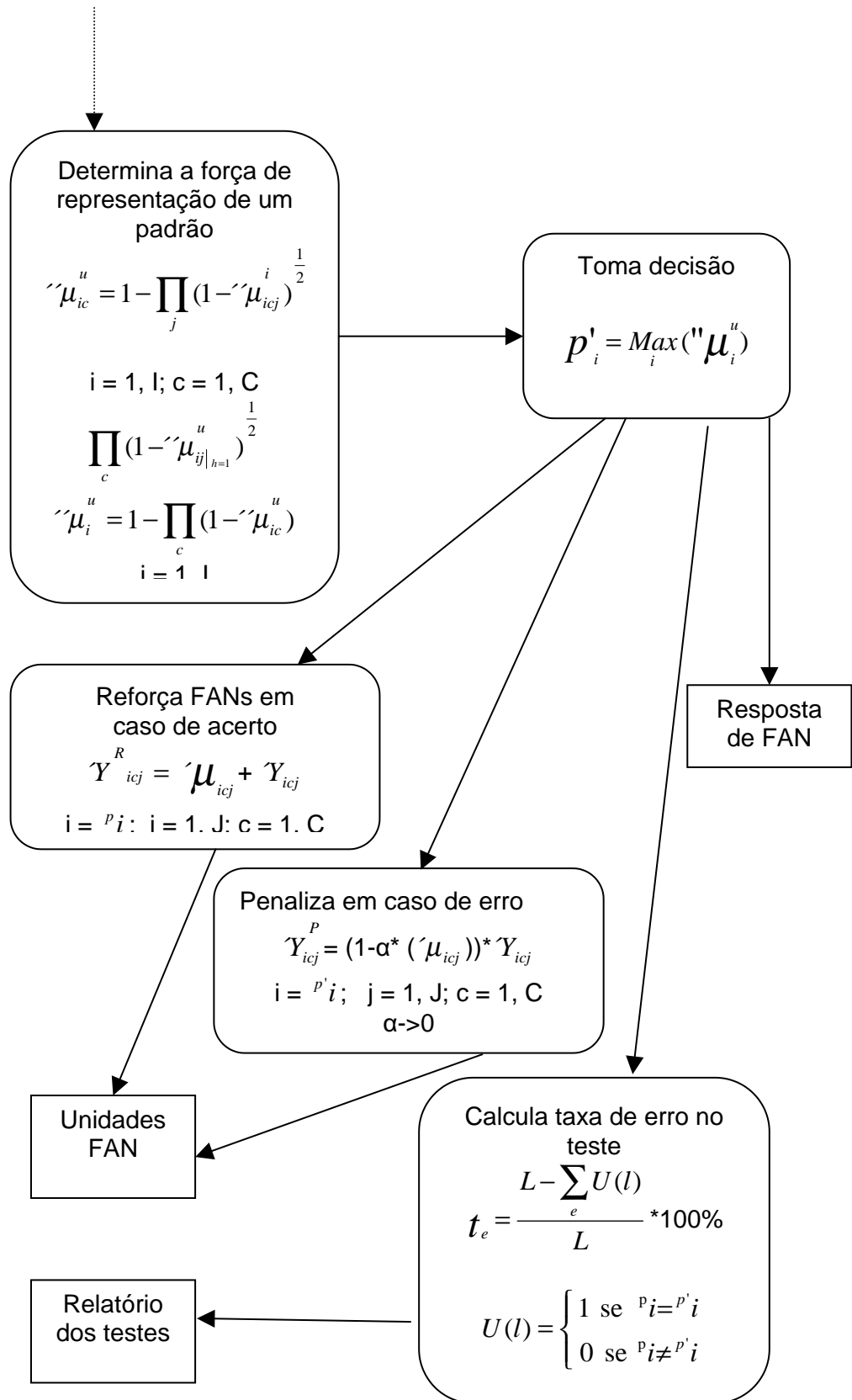


Gráfico 7 – Representação do neurônio correspondente à sétima classe em uma aplicação prática após o treinamento



3.7 Visão esquemática do algoritmo de FAN





No próximo capítulo, são discutidas as aplicações do modelo (FAN) proposto.

4 APLICAÇÕES DE FAN

A primeira aplicação do método FAN para reconhecimento de padrões foi realizada no final do ano de 1996, na solução do conhecido problema de Íris de Fisher (FISHER, 1936).

Tal aplicação pode ser considerada o marco inicial do desenvolvimento do método e pode ser encontrada em RAITTZ (1997) com os comentários e resultados respectivos.

Os bons resultados obtidos na solução desse problema de valor acadêmico apenas motivaram a busca por novos casos de problemas de RP nos quais FAN pudesse ser testado e consistido.

O problema de classificação de cromossomas do grupo de Denver é a mais importante fonte de subsídios para teste e consistência do método de reconhecimento abordado ao longo deste trabalho.

Desde a primeira fase de FAN, até RAITTZ (1997), podem ser observados bons resultados na aplicação do método neste conjunto de dados.

Com as adaptações que são apresentadas nesta tese, no entanto, é que os melhores resultados são obtidos, não apenas em relação a FAN em si mas considerando-se a aplicação de outros métodos sobre o mesmo conjunto de dados.

Por considerar-se esta a aplicação mais importante para a consolidação de FAN, ela é discutida neste capítulo.

Outra aplicação de valor acadêmico de FAN é a sua aplicação na solução do controverso problema do ou exclusivo ou XOR.

Em 1969, uma publicação de MINSKI e PAPERT (1969), abalou a crença depositada nas redes neuronais ao destacar a incapacidade do perceptron em resolvê-lo. A solução desse problema utilizando-se FAN é também discutida neste capítulo.

No que se refere a aplicações de FAN com interesse prático, não apenas acadêmico, comenta-se duas implementações realizadas por outros autores: Mapa FAN no estagiamento automático do sono (DANDOLINI, 2000) e aplicação de FAN na construção de um agente para supervisão de atitudes de alunos em jogos de empresas cujo *software* já esta sendo aplicado, porém não se tem ainda nenhuma publicação, que é esperada para breve.

Os resultados obtidos nas aplicações realizadas até o momento indicam muitos caminhos para aplicações futuras.

4.1 Aplicação no reconhecimento de cromossomas do grupo de Denver

Essa é a aplicação que mostrou que o método discutido abria boas perspectivas para o seu uso em RP. Foi apresentado ao autor deste trabalho e à equipe que pesquisa FAN, por TODESCO (1995), como uma boa forma de testar e avaliar o modelo inicialmente proposto para FAN.

Os testes iniciais mostraram bons resultados e hoje se constata que os melhores resultados conhecidos para tal problema são obtidos nessa abordagem.

Por este trabalho visar mais à demonstração da eficiência de um método de reconhecimento de padrões do que a um estudo sobre cromossomas, discute-se aqui apenas seus aspectos relevantes para RP, deixando de lado

outras questões relativas à origem do problema, detalhadas em TODESCO (1995).

A meta da solução é classificar um cromossoma por meio de suas três características: tamanho, posição do centrômero e densidade de tons de cinza que este apresenta, em uma das sete classes que definem o domínio do problema.

O conjunto de dados disponíveis para treinamento e teste constitui um conjunto de 8111 amostras previamente classificadas de cromossomas, que são divididas em dois subconjuntos, um com 4061 amostras usadas para o treinamento dos neurônios, e outro com 4045 amostras usadas para testar-se a performance do método.

Inúmeros testes e implementações foram realizadas sobre esse conjunto de dados. A seguir são listadas algumas:

1996 - Primeira implementação em Borland-C com parâmetro de combinação de características $H=1$, com resultado obtido sobre o conjunto de testes de 91,0 % de acertos.

1997 - Implementação em Clipper Summer 87, com parâmetro de combinação de características $H=2$, com acertos de 94,6 %, o mesmo obtido por Herington (TODESCO, 1995; RAITTZ, 1997).

1997 - Implementação modificada em Borland-C, com taxa de acerto de 95,25% de acertos.

1997 - Primeira implementação em Delphi visando a obter um produto genérico para o uso de FAN.

1998, 1999, 2000 - Novas alterações no Programa em C com alterações na forma de penalização e eliminação do parâmetro H, em que foi obtido um percentual de acerto superior a 96% (igualando a TODESCO).

2000 - Nova implementação em DELPHI, realizada no laboratório de jogos de empresas da UFSC nos mesmos moldes.

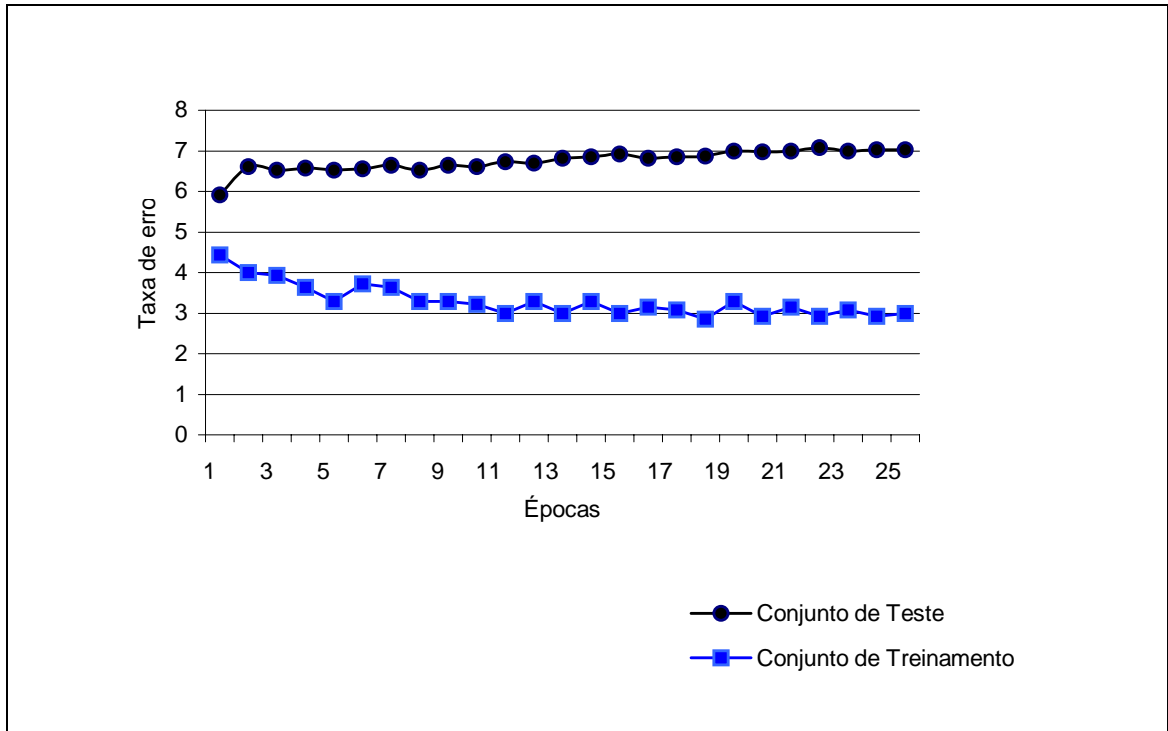
2001 – Obtenção, pela alteração da forma de treinamento, usando-se, para cada época, padrões escolhidos aleatoriamente no conjunto de treinamento com uso de têmpera simulada, a melhor taxa de acertos: 96,36% com erro de 3,64%. A têmpera simulada, nesse contexto, significa a diminuição gradativa da penalização conforme o avanço do número de épocas.

As alterações realizadas em relação às primeiras implementações na forma de treinamento adotada, especialmente no que se refere à penalização, permitem ao método *fugir* de mínimos locais da taxa de erro.

O *Gráfico 8* mostra o comportamento do erro nos estudos de RAITTZ, no ano de 1997 (1).

Pode-se observar que o erro cresce com o número de épocas de forma assintótica tendendo a estabilizar-se no problema dos cromossomas.

Gráfico 8 - Desempenho (Taxa de Erro x Número de Épocas) nos conjuntos de treinamento e teste na aplicação prática dos cromossomas

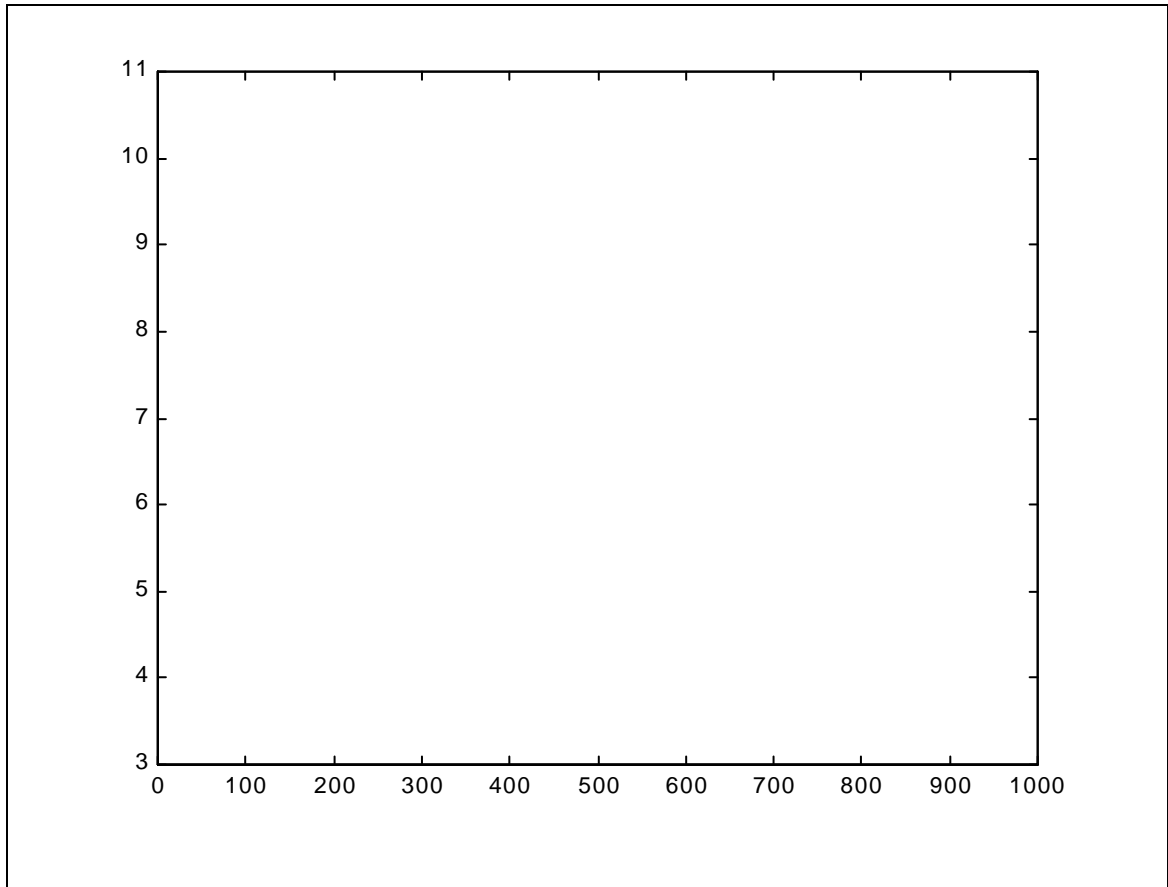


Fonte: Raittz, 1997-1

Na implementação atual, nota-se que a taxa de erro oscila, podendo aumentar por vezes e diminuir por outras. Caso se considere ser possível manter o melhor conjunto de treinamento encontrado em um grande número de épocas, pode-se encontrar o que seja mais adequado para classificação.

Este é o procedimento adotado na obtenção dos resultados descritos na sequência. Abaixo, tem-se o gráfico do número de épocas quando é obtido o melhor resultado de FAN para o problema de classificação de cromossomas no grupo de Denver.

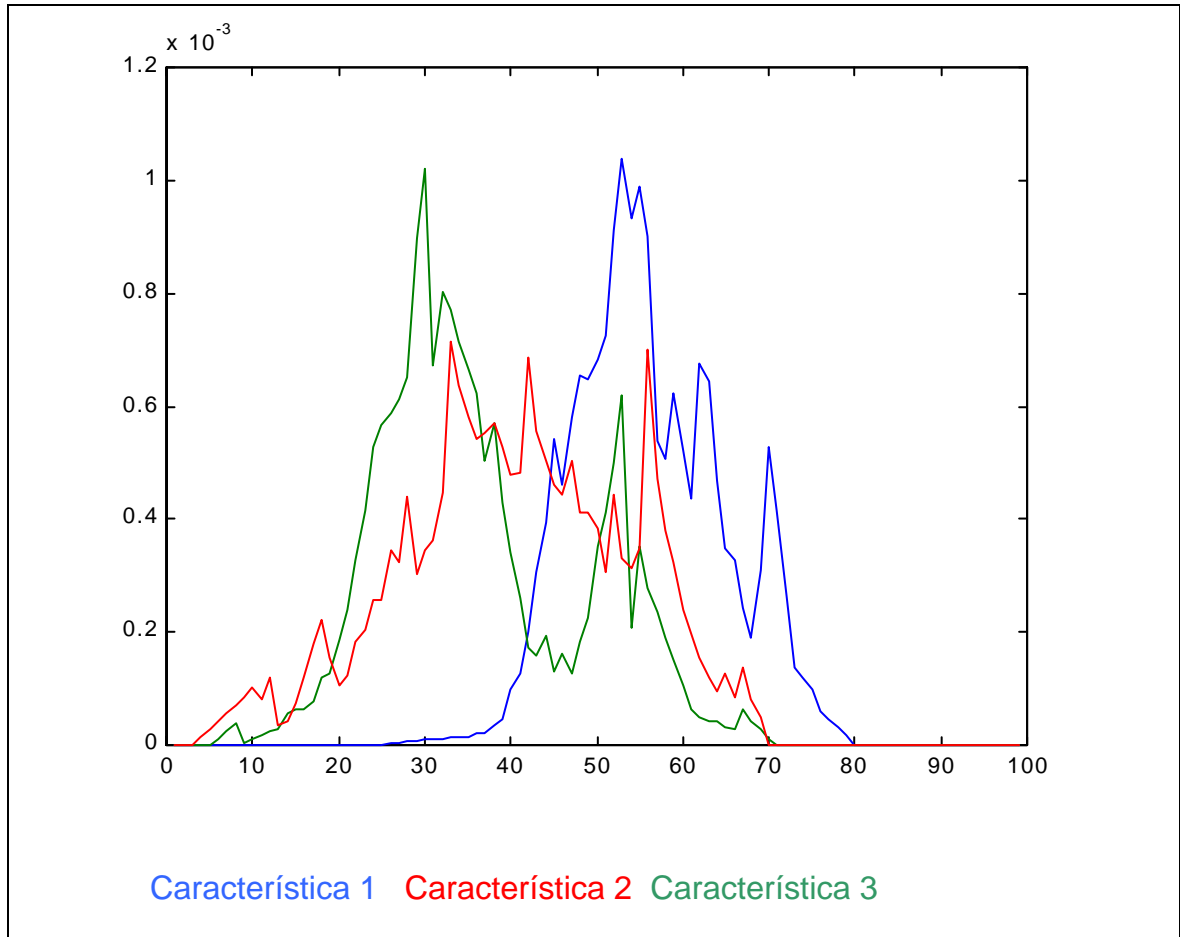
Gráfico 9 - Desempenho (Taxa de Erro x Número de Épocas) no conjunto de teste na aplicação prática dos cromossomas



Na vertical, tem-se a taxa de erro, enquanto na horizontal é indicado o número de épocas realizadas. O ponto onde se encontra o erro mínimo é o que é levado em conta como resultado do treinamento.

Em seguida, é apresentado o gráfico representativo de uma unidade FAN (neurônio) correspondente a esse treinamento. O neurônio é correspondente à terceira classe no grupo de Denver, com a disposição de cada uma de suas características ao longo do eixo dos valores possíveis para as suas ocorrências.

Gráfico 10 – Representação de um dos neurônios na aplicação dos cromossomas do grupo de Denver



Para finalizar a apresentação deste tópico, demonstrar-se-á na tabela abaixo uma comparação entre as taxas de erro em diversas aplicações de técnicas de RP no problema de classificação de cromossomas no grupo de Denver. São apresentados os primeiros resultados obtidos por Herington e Todesco TODESCO (1995); e os resultados obtidos por RAITTZ por meio de FAN, primeiro em 1997, e atualmente (2002).

Tabela 3 – Comparação entre as taxas de erro em diversas aplicações no problema dos cromossomas do grupo de Denver

MODELO	TESTE (%)
Multi-Layer Perceptron Backpropagation (Herington)	5,40
Radial Basis Functions (Todesco)	4,04
Free Associative Neurons 1997	4,75
Free Associative Neurons 2002	3,64

4.2 Aplicação na resolução do problema do XOR

Como já foi comentado, a não obtenção de uma solução para o problema do XOR pelos primeiros perceptrons causou uma grande polêmica relativa à validade do uso das redes neuronais.

A questão surge novamente ao se tentar justificar o uso de um novo método que possua afinidades conceituais com estas redes. A questão é: é possível resolvê-lo com a proposta apresentada em FAN? Vários testes foram realizados.

Buscar resolver o problema do XOR com uma técnica de RP demanda certas definições. Pode-se compreendê-lo como um problema com duas classes possíveis $c \in \{0,1\}$ e com duas características x que podem assumir valores 0 ou 1, conforme a *Tabela 4*:

Tabela 4 – Padrões derivados do problema do XOR

x_1	x_2	C
0	0	0
0	1	1
1	0	1
1	1	0

Tem-se assim um conjunto de treinamento preparado para um treinamento supervisionado aplicável aos modelos definidos de FAN. Pode-se realizar o teste utilizando-se o mesmo conjunto, uma vez que os dados são limitados pela definição do problema. Resta então aplicar a técnica e verificar os resultados.

Constata-se que ao se aplicar a primeira abordagem (RAITZ, 1997) com parâmetro de combinação de características $H=1$ neste problema, esta falha não atinge 100% de acertos. O mesmo não ocorre se aumentar tal parâmetro, no caso $H=2$, quando não ocorre erro algum de classificação. Portanto, a primeira abordagem de FAN consegue resolver o problema do XOR, desde que haja um parâmetro de combinação de características $H>1$.

Já comentou-se que na definição atual do modelo FAN foi abandonado o uso do Parâmetro H . Isto sugere que este não possa obter sucesso ao se classificarem padrões com características semelhantes às encontradas no XOR. Ocorre que essas características não são independentes entre si, o que a estrutura do algoritmo de FAN prevê a princípio.

Para se tratar problemas como este, evitando o gasto computacional que o uso do parâmetro H implica, definiu-se que a combinação das características pode ser substituída pela inclusão de uma ou mais características adicionais que represente a combinação desejada nos problemas nos quais isso é necessário.

Elaborou-se uma técnica que suprime a demanda pela combinação entre características nos moldes como essa ocorre em FAN-97, que permite a resolução do XOR e de outros problemas similares de forma genérica, não afetando a solução de outros problemas.

Consiste no seguinte: definir uma transformação matemática que leve o espaço do conjunto características a serem combinadas ao espaço unidimensional gerando, assim, uma nova característica a ser usada conjuntamente com as demais.

Pode-se fazê-lo usando a técnica mais adequada de redução de dimensionalidade, a ser definida para o problema (SOUZA, 1999). Incluída a nova característica, a combinação é levada em conta em FAN, sem que ocorra a sobrecarga computacional implicada pelo método antigo.

A expressão utilizada para a geração da terceira característica extra X_3 no caso em questão (XOR) é definida pela fórmula:

$$X_3 = \sqrt{(x_1 * x_2)/2}$$

Com a inclusão dessa característica, o problema é resolvido com facilidade em poucas épocas de treinamento. Os *Gráficos 11 e 12* mostram a situação dos neurônios treinados com sucesso na solução do problema do XOR com as duas características originais, mais a terceira incluída pelo método.

Gráfico 11 – Representação do primeiro neurônio correspondente à primeira classe no problema do XOR, após o treinamento

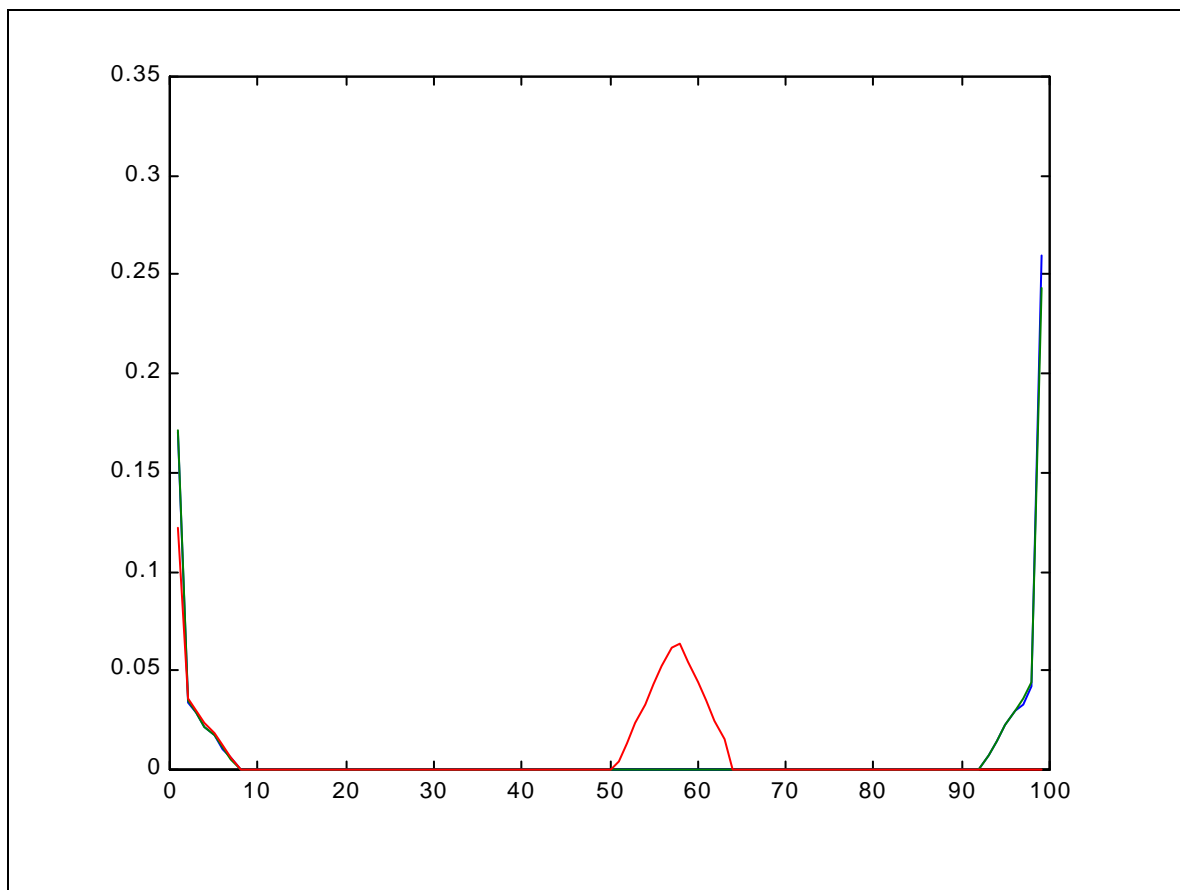
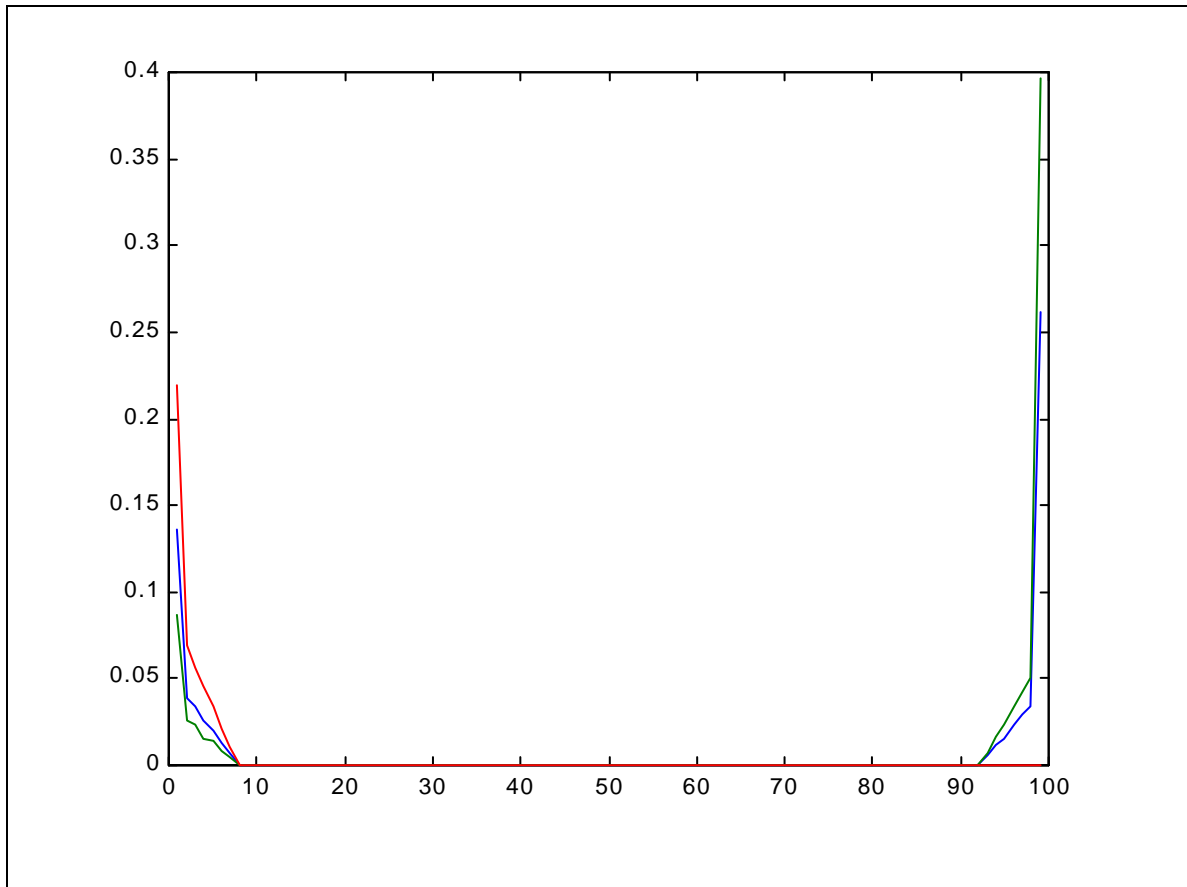


Gráfico 12 – Representação do segundo neurônio correspondente à segunda classe no problema do XOR, após o treinamento



Nos dois gráficos tem-se em azul e verde as características originais; em vermelho, é apresentada a característica extra.

Vale comentar que se espera que esta técnica de inclusão de características extras possa ser eficaz na resolução do problema do XOR usando-se outras técnicas como o próprio perceptron.

A consideração quanto a isso é que os problemas de reconhecimento devem ser redefinidos quando ocorre a dependência entre as características.

Acrescenta-se, no entanto, que em qualquer implementação de FAN, quer haja dependência ou não entre características, uma nova característica extra pode ser incluída como resumo das demais sem afetar os resultados, e de

forma genérica como uma estratégia do método sem afetar de forma significativa os resultados.

A mesma característica extra usada no XOR, por exemplo, foi usada em um teste no problema dos cromossomas (tópico anterior) e a taxa de erro no melhor resultado obtido foi inferior a 4%. Isto implica na necessidade de inclusão de uma característica adicional, representante do conjunto das características no modelo, com o objetivo de torná-lo mais genérico.

4.3 Mapa FAN

O mapa FAN é uma técnica de RP derivada de FAN (conforme DANDOLINI 2000). A primeira apresentação desta técnica ocorreu no ano de 2000 pela mesma autora em virtude de uma aplicação no problema de reconhecimento das diversas fases do sono.

Tal aplicação é comentada neste trabalho devido à sua íntima relação com o método tratado e por ilustrar a capacidade de serem utilizados neurônios independentes para cada classe na solução deste tipo particular de problema.

Aqui apresenta-se o Mapa FAN e sua aplicação de forma resumida, podendo os interessados obterem outras informações nos referenciais de DANDOLINI (2000).

De acordo com a autora, o objetivo do Mapa FAN é *“(...) tornar possível a sua utilização em problemas multidimensionais (como em reconhecimento de voz, em classificação de imagens de satélite e na classificação dos estágios do sono)”*.

Nesse tipo de problema, as características dos padrões relevantes para a classificação não se encontram de forma explícita como nos exemplos dos casos já vistos. Tem-se, na verdade, um grande volume de características para cada padrão definido por uma sequência de valores de amplitudes de sinais gerando padrões de grande dimensão.

A abordagem convencional de FAN não seria adequada para o tratamento desse tipo de problema se usada diretamente sobre as informações dos padrões.

Para se usar a abordagem convencional de FAN, deve-se extrair, a partir dos padrões multidimensionais desses problemas, características que os resumam como médias, desvios padrão e transformações matemáticas, e só então utilizar-se estes padrões modificados para a classificação.

Os Mapas FAN se propõem a resolver tais problemas de forma genérica, sem a necessidade da transformação das características originais.

Prossegue DANDOLINI (2000), em sua explicação:

“Os Mapas FAN são matrizes bidimensionais ($d \times n$), onde uma dimensão se refere às amplitudes dos sinais e a outra à ordem do sinal amostrado. Para cada classe é criado um Mapa FAN. Cada padrão de entrada é apresentado ao Mapa FAN que representa a classe do padrão. Desta forma, tem-se um processo de aprendizagem supervisionado. O objetivo é modelar cada mapa com aspectos da classe que ele representa.

Um padrão a ser testado é apresentado a cada mapa treinado e este retorna o grau de pertinência daquele padrão em relação à classe que aquele mapa representa. O padrão de entrada será classificado de acordo com o maior grau de similaridade (pertinência)”.

4.3.1 A aplicação no estagiamento automático do sono

O problema do estagiamento automático das fases do sono pode ser considerado como um problema de reconhecimento de padrões supervisionado.

Existem várias fases do sono que são focos de interesse dos especialistas. Estagiar o sono consiste basicamente identificar cada uma dessas fases em função das informações obtidas nos canais de eletro (eletrodos) conectados aos pacientes.

Uma vez que se pode dispor de exames já realizados e avaliados por médicos especialistas, é possível modelar padrões para treinamento de sistemas de RP, os quais, uma vez treinados, permitem o seu uso para a realização de outros estagiamentos automáticos, diminuindo o trabalho dos especialistas.

Acerca do estagiamento do sono, DANDOLINI (2000) discorre que

“A polissonografia mostra uma sucessão de ondas cerebrais ordenadas e cíclicas de diferentes amplitudes e frequências, de movimentos oculares e de mudanças de tono muscular que são utilizados para o estagiamento do sono. A seqüência das fases, a precisa cronologia dos eventos, foi denominada arquitetura do sono. A arquitetura de uma noite de sono é representada através dos hipnogramas.

O estagiamento é realizado pela análise de no mínimo três canais polissonográficos e no máximo o limite do aparelho. Os sistemas computadorizados, hoje em uso, suprimam o uso do papel e registram com facilidade 30, 60 ou mais canais, caso o pesquisador necessite. O computador auxilia, também, no estagiamento, porém, ainda sem grande precisão (Martinez, 1999).

O estagiamento do registro do sono é realizado, normalmente, por conveniência, em segmentos chamados de épocas. O tamanho médio

desses segmentos é de 15 a 40 segundos de amostra, que correspondem, aproximadamente, a uma página de papel de 30 cm registrado a uma velocidade de 10 a 15 milímetros por segundo. O estagiamento de uma época inferior a 20 segundos é considerado extremamente tedioso por muitos grupos, embora existam grupos que cheguem a utilizar épocas de até 3 segundos.

A aparência dos canais muda a cada instante. No curto período de uma época de 30 segundos, pode-se passar de vigília para estágio 1, para estágio 2 e retornar à vigília. A época recebe o rótulo da fase que predominar.

Outros tipos de medidas, como o caso da turgescência peniana durante o sono REM, conforme observado por Karacan e Shapiro (1966), também podem ser utilizados para auxiliar na codificação da fase da época em análise. Outros fenômenos, como a respiração e a saturação do oxigênio, também são apontados por Aserink e Kleitman (1953). Outros dados, anotados pelo técnico que acompanha o sono do paciente, também são utilizados no momento do estagiamento (Martinez, 1999)".

Nessa aplicação, o sinal digital dos eletroencefalogramas-EEG são subdivididos em intervalos de tempo com uma seqüência significativa de amplitudes, denominadas de épocas, que são, por sua vez, ordenados (RAITZ et al., 1999). Ordenado o sinal, obtém-se uma transformação do padrão com melhores condições de comparação com outras épocas permitindo obter generalidade na classificação.

O sinal ordenado é usado para treinar um determinado neurônio ao qual o padrão está relacionado no caso um mapa FAN que representa uma das classes em questão (estágio do sono específico); ou, se for o caso, testar a sua similaridade com cada um dos neurônios do modelo para escolher a que melhor o representa.

Um sistema que se utiliza dos Mapas FAN conjuntamente com outras técnicas de Inteligência artificial foi desenvolvido e está apresentado em

detalhes por DANDOLINI (2000). A seguir, acrescentar-se-ão comentários da autora acerca da utilização do mapa FAN no auxílio ao estagiamento automático do sono para finalizar este tópico.

“(...) Outro fator importante é que a ordenação (pré-processamento) e o Mapa FAN (classificador) que compõem o modelo, apesar de serem métodos novos, demonstraram resultados comparáveis a outros métodos, o que motiva uma continuação das pesquisas. A ordenação poderá ser utilizada como parte do pré-processamento de sinais digitais para outros classificadores de padrões de sinais digitais amostrados temporalmente, ou seja, dependentes de frequência e tempo de cada amostra.”

4.4 Utilização de FAN para a construção de um agente para supervisão de alunos em jogos de empresas

No ano de 1999, surgiu a demanda nesta universidade por uma técnica de RP a ser aplicada em um modelo de agente para facilitar aos professores supervisão e orientação aos alunos de jogos de empresa.

O problema dos alunos, de acordo os professores, era o de que houvesse coerência por parte deles na definição da demanda que sua empresa imaginária provocara, dadas as suas atitudes no jogo.

Por exemplo: um aluno poderia definir um preço muito baixo para o seu produto sem que possuísse estoque suficiente para suprir a respectiva demanda. Então, cabia ao professor, em tal caso, orientar ou estimular o aluno a pensar se a jogada condizia com seus objetivos no jogo.

Além do preço, outras variáveis afetariam a demanda e, também, as jogadas dos outros jogadores e o cenário definido pelo ambiente do jogo poderiam interferir nos resultados.

Ainda conforme um dos professores, a definição de regras para o estabelecimento de eurísticas no intuito de prover o agente com informações acerca da demanda era difícil e os resultados do uso desta estratégia, imprevisível.

A questão levantada fora a seguinte: *“é possível que uma estratégia de FAN de RP resolva o problema do reconhecimento do nível de demanda a ser esperado dada a jogada de um aluno?”*.

Foi o que se buscou responder durante mais ou menos seis meses de trabalho conjunto, na perspectiva de FAN/Jogos de Empresas.

O laboratório de Jogos de Empresas, dispõe de milhares de jogadas realizadas por alunos em aulas reais, em diferentes instituições e com variáveis de contexto semelhantes.

O RP para a construção do agente foi viabilizado ao se extraírem características relevantes desses dados e a definição das classes de interesses no domínio do problema em questão.

Optou-se por definir três características principais para o desenvolvimento de um sistema piloto para ser testado. Tais características foram definidas pela jogada do aluno, que eram as variáveis a serem definidas no intervalo $[0,1]$ conforme a seguir:

- o preço definido pelo aluno;
- o nível de propaganda estipulado;
- o prazo dado para a venda.

Por sua vez, as classes definidas no domínio do problema eram:

- **classe 1:** demanda alta – as que apareciam no conjunto das 25% das melhores demandas obtidas por alunos em todos os dados do conjunto definido para o treinamento da rede;
- **classe 2:** demanda baixa – demandas encontradas no conjunto das 75% das melhores demandas e que não eram demanda alta;
- **classe 3:** demanda baixa (todas as demais).

Com tais definições e os dados em mãos, aplicou-se FAN em uma das implementações já existentes para verificar os resultados.

Os testes obtidos na experiência piloto mostram o sucesso do método ao realizar a classificação das demandas no problema proposto, atendendo às necessidades dos professores.

A partir dessa fase, o agente foi construído tendo nele embutido um sistema de neurônios FAN para a classificação de padrões apto a abarcar novas características, conforme a necessidade destas serem usadas. Estando o agente pronto, ele é incluído no sistema de jogos de empresas daquele grupo de trabalho.

Vale destacar que, com a aplicação de FAN, os jogos com o agente são mais estimulantes e os alunos se sentem mais seguros.

A avaliação dos professores é que *“a experiência com o uso de FAN em agentes atende à expectativa”*.

5 CONCLUSÃO

Quanto à publicação de FAN sobre reconhecimento de padrões (RAITZ 1997), verifica-se uma nova abordagem, com menos de um ano de existência.

Nela, são expostos os conceitos básicos do modelo e algumas aplicações que buscam mostrar o potencial do novo método de RP.

Muita coisa sucedeu desde então: conceitos de treinamento foram modificados, melhorando a performance do método; uma nova descrição matemática – mais clara – foi definida; novos resultados foram obtidos em aplicações já realizadas; novas publicações e apresentações do método foram feitas e, acima de tudo, FAN já está sendo utilizado em problemas práticos.

Tal percurso justifica esta publicação.

Vale destacar que a demanda por técnicas de uso geral de RP ocorre sobretudo nos meios técnicos e científicos. Biólogos precisam de ferramentas de classificação de espécimes; programadores buscam formas eficazes de aprendizado para realizar tomadas de decisões baseadas em dados, e profissionais de diversas outras áreas, cuja afinidade com a inteligência artificial e, mais especificamente com o reconhecimento de padrões não é tão grande, vivenciam problemas cuja solução, por meio da aplicação de RP, é adequada ou necessária.

A construção da ferramenta apresentada neste trabalho, associada ao conhecimento de causa que esta proporciona aos seus autores pelo percurso que realizaram ao construí-la, deve vir a facilitar o intercâmbio entre essas áreas do conhecimento e a forma de solucionar os problemas de RP na comunidade acadêmica brasileira.

A pertinência do método aqui apresentado pode ser verificada ao se levarem em conta as alterações produzidas neste trabalho, em relação aos demais publicados a respeito de FAN. O algoritmo apresentado no capítulo 3 fornece os subsídios necessários à implementação computacional do modelo FAN 2002. Além disso este trabalho permite ao leitor interessado em RP acompanhar o desenvolvimento histórico do modelo ao longo de vários anos.

A diferença entre FAN 1997 e FAN 2002 pode ser avaliada se forem comparados os resultados obtidos no caso de estudo dos cromossomas do grupo de Denver.

Na primeira, a taxa de erro obtida era de 4,75%, enquanto na segunda a taxa de erro é de 3,64%. Torna-se procedente registrar que a menor encontrada em todas as publicações para este problema é 23% menor que a taxa obtida no mesmo método anteriormente.

De fato, pode-se observar a evolução do modelo comparativamente às publicações já realizadas do modelo anterior, inferiores ao atual, e as publicações subseqüentes, em que constam já as primeiras alterações do modelo FAN 2002 (RAITZ 1997-1, RAITZ 1997-2, RAITZ 1998).

O modelo se destaca porque já subsidia o desenvolvimento de outras técnicas de RP, com várias implementações genéricas de classificação de padrões, nas quais um usuário de outra área do conhecimento pode encontrar a forma para solucionar problemas sem um domínio completo das técnicas específicas de reconhecimento de padrões.

Casos como a aplicação em jogos de empresas acontecem com frequência no meio acadêmico brasileiro. A opção pelo uso de FAN nesta

aplicação foi marcantemente definida pela possibilidade do usuário interessado poder interagir com o desenvolvedor do método.

Um programa para uso geral de FAN 2002 foi desenvolvido e hoje, os resultados dessa aplicação já fazem parte da realidade de outros ambientes de estudos independentemente da área específica de RP.

No futuro, pretende-se continuar a divulgação e facilitação da utilização do método. Para tal, deve-se pôr em disponibilidade uma implementação de uso geral, nos moldes apresentados neste trabalho.

Ainda, está prevista a inclusão de FAN como técnica de RP nos conteúdos da disciplina de *Inteligência Aplicada*, no curso de Tecnologia em Processamento de Dados da Escola Técnica da Universidade Federal do Paraná, com o objetivo de capacitar os alunos a resolverem problemas de RP e dar viabilidade ao desenvolvimento ulterior de FAN em um ambiente que ultrapasse aquele restrito a seus autores.

Como sugestão para trabalhos futuros pode-se indicar a necessidade de um estudo aprofundado da métrica utilizada para os operadores difusos usada no modelo buscando a sua justificativa e a construção de bases de conhecimento FAN para serem utilizadas conjuntamente com bancos de dados.

Espera-se que o estabelecimento desse método original e genérico para a classificação de padrões sirva sobretudo aos interessados em sua aplicação e, ainda, que ele possa encorajar outras iniciativas de busca de novas técnicas de RP e de *inteligência artificial* em nosso país.

FONTES BIBLIOGRÁFICAS

AUSTIN J. AND LEES K. ***A Search Engine based on Neural Correlation Matrix Memories***. *Neurocomputing*. Vol. 35. pp. 55-72, 2000.

BASTOS, R. C. ***Avaliação de desempenhos educacionais: uma abordagem usando conjuntos difusos***. Tese de Doutorado. Universidade Federal de Santa Catarina. Departamento de Engenharia de Produção, 1994.

BISCHOP, C. M. ***Neural Networks for Pattern Recognition***. Oxford University press, 1995.

CICHOCKI A.; UNBENHABEN, R. ***Neural Networks for Optimization and Signal Processing***. John Wiley & Sons inc, 1993.

DANDOLINI, G. A. ***Um procedimento para avaliação da saúde financeira de pequenas empresas: estudo de um caso usando redes neuronais artificiais***. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção, 1996.

DANDOLINI, G. A. ***Mapa FAN no estagiamento automático do sono***. Tese de Doutorado, Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção, 2000.

D'ALVEYDRE, S. Y. ***El arqueómetro: clave de todas las religiones y todas las ciencias de la antigüedad***. Luis Carmo, editor, Madrid. 1997.

DORON, R. & PAROT, F. ***Dicionário de psicologia***. Editora Ática, 1998.

DURKIN, J. ***Expert Systems: Design and Development***. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1994.

FISHER, R.A. ***The Use of Multiple Measurements in Axonomic Problems***. 1936. [doc WWW Data: 15/10/1997].
<http://www.comcat.com/~hutch/DASL/dataex.htm>

FU, L. M. ***Rule Generation from Neural Networks***. *IEEE Transactions on Systems. Man and Cybernetics*. Vol. 28. No. 8. pp. 1114 - 1124, 1994.

FUKUSHIMA, K.; MIYAKE, S.; ITO, T. ***Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition***. *IEEE Transaction on Systems. Man, and Cyber tics* SMC-13. pp. 826-834, 1983.

GOLDBERG, D. E. ***Genetic Algorithms in Search, Optimization, and Machine Learning***. Addison-Wesley, 1989.

GOONATILAKE, S.; KHEBBAL, S. ***Intelligent Hybrid Systems. Proceedings of First International Conference on Intelligent Systems***. Singapore. 1992.

GOONATILAKE, S.; Khebbal, S. **Intelligent Hybrid Systems: Issues, Classifications, and Future Directions.** in [Goonatilake & Khebbal, 1995a].

GROHMAN W. M. AND DHAWAN A. P. **Fuzzy convex set-based classification for analysis of mammographic microcalcification.** *Pattern Recognition*. Vol. 34. pp. 1469-1482, 2001.

HAYKIN, S. **Neural Networks: A comprehensive Foundation.** Macmillan College Publish Company. New York, 1994.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis.** Prentice Hall Inc., 1982.

KANDEL, A. e LANGHOLZ, G., **Hybrid Architectures for Intelligent Systems,** Ed. CRC Press. Inc., 1992.

KLIR, G. J. & YUAN, B. **Fuzzy sets and fuzzy logic - theory and applications.** Prentice Hall, 1995.

KOSKO, B. **Neural Networks and Fuzzy Systems.** Prentice Hall, Englewood Cliffs, NJ, 1992.

LAPOLLI, E. M. **Processamento de imagens digitais: uma abordagem utilizando conjuntos difusos.** Tese de Doutorado, Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção, 1994.

LEIGH W. AND PURVIS R. AND RAGUSA J. M. **Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support.** *Decision Support Systems*. Vol. 32. pp. 361-377, 2002.

LEV, B., and SUNDER, S. **Methodological issues in the use of financial ratios.** *Journal of Accounting and Economics* 1/3,(1979) 187-210.

LEV, B. **Financial Statement Analysis,** Prentice-Hall, Englewood Cliffs, NJ, 1974.

MARTINS, A. R. **Modelling Financial Statements Analysis for Indicating Corrective Actions to Financial Problems.** Tese de Doutorado, Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção, 1996.

MARTINS, A.; PACHECO, R.; WEBER-LEE, R.; BARCIA, R. M. **Integrating Expert Networks and CBR in a Hybrid Architecture to Analyze the Financial Health of a Firm,** Brazil 2nd International Congress of Industrial Engineering and 16th National Congress of Production Engineering - ENEGEP , October, 7 - 10, Piracicaba, São Paulo, Brasil, 1996.

MCNEILL, D., FREIBERGER P. **Fuzzy Logic: The Revolutionary Technology that is Changing our World.** Touchstone, 1993.

MEBES, G. O. Os arcanos maiores do tarô: curso de enciclopédia do ocultismo. Editora Pensamento, São Paulo, 1998.

MEDSKER, L. **Híbrido Inteligente Sistemas**. Boston: Kluwer Academic Publishers, 1995.

MEYER, P. L. **Probabilidade: aplicações à estatística**. Rio de Janeiro. Livros Técnicos e Científicos Editora Ltda. Rio de Janeiro, 1975.

MINSKI, M.; PAPERT, S. **Perceptrons**. Cambridge, MA. MIT Press, 1969.

MOSHOU ET AL. **A neural network based plant classifier**. *Computers and Electronics in Agriculture*. Vol. 31. pp. 5-16, 2001.

PACHECO, R. C. S. **A Hybrid Intelligent System for Diagnosing and Solving Financial Problems of Small Retail Firms**. Ph.D. Dissertation, Universidade Federal de Santa Catarina UFSC, Programa de Pós-graduação em Engenharia de Produção, Brasil, 1996.

QUEK C. AND TUNG W. L. **A novel approach to derivation of fuzzy membership functions using the Falcon-MART architecture**. *Pattern Recognition Letters*. Vol. 22. pp. 941-358, 2001.

RAMOS, P. G. **Uma investigação das redes neuro-fuzzy aplicadas à mineração de dados**. 1999. [doc WWW Data: 16/05/2002].
<http://www.di.ufpe.br/~tg/1999-1/p>

RAITZ, R. **Free Associative Neurons – FAN: uma abordagem para reconhecimento de padrões**. Florianópolis, 1997. Dissertação de Mestrado em Engenharia de Produção. Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina.

RAITZ, R. T.; SOUZA, J.A.; DANDOLINI, G. A.; PACHECO, R. C. S.; MARTINS, A., GAUTHIER, F & BARCIA, R. **Learning by Means of Free Associative Neurons**. NAFIPS 97: *Annual Meeting Of The North American Fuzzy Information Processing Society*, September 21-24, 1997, Syracuse, New York.

RAITZ, R. T., SOUZA, J. A. DANDOLINI, G. A., et al. **FAN: Learning by Means of Free Associative Neurons**, WCCI98 - IEEE World Congress on Computational Intelligence, Anchorage Alaska, pp. 425-430, May 1998.

RAITZ, R. T., SOUZA, J. A. DANDOLINI, G. A., et al. **Pré-processamento de sinais digitais**. XIX ENEGEP - Encontro Nacional de Engenharia de Produção, 1999.

ROSS, T. J. **Fuzzy Logic with Engineering Applications**. McGraw-Hill Inc., 1995.

ROY, D. K. AND PENTLAND A. P. ***Learning Words from Sights and Sounds: a Computational Model***. *Cognitive Science*. Vol. 26. pp. 113-146, 2002.

RUIZ, J. A. **Metodologia científica: guia para a eficiência nos estudos**. Editora Atlas, SA, 1986.

SCHALKOFF, R. J. ***Pattern Recognition: Statistical, Structural and Neural Approches***, John Wiley & Sons, Inc., 1992.

SHAMS ET AL. ***Graph matching vs mutual information maxization for objects detection***. *Neural Networks*. Vol. 14. pp. 345-354, 2001.

SHROUT, R. N. **Hipnose científica moderna: dos mistérios da antigüidade à ciência contemporânea**, Pensamento, 1985.

SCHÜGERL, K. ***Progress in monitoring, modeling and control of bioprocesses during the last 20 years***. *Journal of Biotechnology*. Vol. 85. pp 149 – 173, 2001.

SOUZA, J. A. **Reconhecimento de padrões usando indexação recursiva**. Florianópolis, 1999. Tese de Doutorado em Engenharia de Produção. Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina.

TAFNER, M. A. **Estagiamento automático do sono utilizando um canal de EEG e uma rede neuronal artificial com alta representação cortical**. Florianópolis, 1999. Tese de Doutorado em Engenharia de Produção. Programa de Pós-graduação em Engenharia de Produção, Universidade Federal de Santa Catarina.

TODESCO, José L. **Reconhecimento de padrões usando rede neuronal artificial com uma função de base radial: uma aplicação na classificação de cromossomas humanos**. Tese de Doutorado, Universidade Federal de Santa Catarina, Departamento de Engenharia de Produção, 1995.

WAKAHARA T. AND KIMURA Y. ***Toward robust handwritten Kanji character recognition***. *Pattern Recognition Letters*. Vol. 20. pp. 979-990, 1999.

ZADEH, L.A. ***Fuzzy sets. Information and Control***, Vol. 8, 1965, pp. 338-353.

ZADEH, L.A. ***Fuzzy algorithms, Information & Control***, Vol. 12, 1968, pp. 94-102.

ZADEH, L.A. ***Making computers think like people***, IEEE Spectrum 8, pp. 26-32, 1984.

ZHOU ET AL. ***Lung cancer cell identification based on artificial neural network ensembles***. *Artificial Inteligence in Medicine*. Vol. 24. pp. 25-36, 2002.